

TITLE National Curriculum: Task Group on Assessment and Testing. A Report.

INSTITUTION Department of Education and Science, London (England).; London Univ. (England). Centre for Educational Studies.; Welsh Office, London (England). Education Dept.

PUB DATE Dec 87

NOTE 198p.

PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC08 Plus Postage.

DESCRIPTORS Curriculum Development; *Curriculum Evaluation; *Educational Assessment; Elementary Secondary Education; Foreign Countries; Measurement Techniques; *National Programs; Program Descriptions; *Student Evaluation; *Test Use

IDENTIFIERS *Great Britain; Task Force Approach

ABSTRACT

Assessment and testing associated with implementation of the national curriculum developed by the British Department of Education and Science (London) are reviewed. Topics addressed include: profile components in the assessment framework, assessment and testing methods, bias in assessments, assurance of confidence in national assessment, reporting, the relation of criteria and scales to expected routes of child development, implementation of the curriculum in primary schools and secondary schools, assessment of children with special needs, the work of the national curriculum subject working groups, phasing of the curriculum implementation, the support structure, teacher training, educational materials, the role of the local educational authority, and future work for the task group designated to advise on assessment and testing within the national curriculum. Appendices include terms of reference and membership in the task group; a letter of guidance from the British Secretary of State; an annotated list of meetings; examples of test and assessment methods available; samples of tasks for students; a paper by the Equal Opportunities Commission ("National Testing and Equal Opportunities"); a technical note on reliability, validity, and error; guidelines for the conduct of group moderation meetings; records of achievement; a technical note on background factors affecting educational performance; school enrollment statistics; outlines of how the system can be implemented in primary schools and secondary schools; and data on existing and planned provisions for resource needs. (TJH)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

National Curriculum

Task Group on Assessment and Testing

A Report

BEST COPY AVAILABLE

Department of Education and Science
and the Welsh Office

National Curriculum

Task Group on Assessment and Testing

A Report

Department of Education and Science
and the Welsh Office

REPORT OF THE TASK GROUP ON ASSESSMENT AND TESTING

CONTENTS

Preface and Glossary

Letter to the Secretary of State

Sections of Main Report:

Paragraphs:

INTRODUCTION

I.	Our Starting Point	1 - 6
II.	Present Practice	7 - 12
III.	Concerns to be Addressed	13 - 19
IV.	The Structure of the Report	20 - 22

PURPOSES AND PRINCIPLES

V.	Purposes	23 - 30
VI.	Profile Components in the Assessment Framework	31 - 39
VII.	Assessment and Testing Methods	40 - 50
VIII.	Bias in Assessments	51 - 53
IX.	Ensuring Confidence in National Assessment	54 - 63
X.	Moderation	64 - 90

THE ASSESSMENT SYSTEM IN PRACTICE

XI.	Progression, Scales and Ages	91 - 124
XII.	Reporting	125 - 138
XIII.	The Primary Stage	139 - 157
XIV.	The Secondary Stage	158 - 167
XV.	The Assessment of Children with Special Educational Needs	168 - 173

IMPLEMENTATION

XVI.	The Brief for the Subject Groups	174 - 188
XVII.	Phasing	189 - 199

A SYSTEM OF SUPPORT

XVIII.	The Support Structure	200 - 205
XIX.	Inset and Other Resource Needs	206 - 214
XX.	The Local Education Authority's Role	215 - 217
XXI.	Future Work for the Task Group	218

SUMMARY AND CONCLUSIONS

XXII.	Conclusions	219 - 226
XXIII.	A List of Recommendations	227

PREFACE AND GLOSSARY

This report of the Task Group on Assessment and Testing ventures into a specialised field with its own technical vocabulary. We have tried to avoid technical jargon wherever possible, but it has sometimes been necessary to use terms which are unfamiliar to the layman or whose meaning here is different from that of everyday usage. We explain some of these terms below. Others are explained where they first appear in the report - including those we have specially coined to label ideas newly devised by the Group.

Assessment

A general term encompassing all methods customarily used to appraise performance of an individual pupil or a group. It may refer to a broad appraisal including many sources of evidence and many aspects of a pupil's knowledge, understanding, skills and attitudes; or to a particular occasion or instrument. An assessment instrument may be any method or procedure, formal or informal, for producing information about pupils: eg. a written test paper, an interview schedule, a measurement task using equipment, a class quiz.

Test

Strictly, any assessment conducted within formal and specified procedures, designed to ensure comparability of results between different test administrators and between different test occasions. For some it implies a set of written questions, externally prescribed, with written responses marked according to rigid rules; for others, any of a broad range of assessment instruments with standardised rules of administration and marking which ensure comparability of results. This report uses the term in the latter, broader sense. Tests intended for widespread use are commonly pilot-tested with a representative national sample. A standardised test is one published with marking instructions and data on the levels of response established for a national population sample, at a given age, or with different data for several ages.

Tasks and projects

A task is any activity in which a pupil takes part and which could be used for assessment. Examples are a written test, an essay, a practical job such as designing or constructing an artefact, or conducting an experiment. Tasks which are open-ended and extensive in scope and/or in time are often called projects: these may require independent planning and/or research on the part of the pupil.

Item bank

An orderly collection of test questions or tasks, so assembled and labelled that users may select from it sets of tasks for particular assessment purposes. Those in charge of such a bank usually exert quality control on the questions put in it.

Ranking, rating and marking

Ranking involves putting pupils' performances in order of relative quality. The set of rankings forms a scale. Where a scale simply has some order, it may be called a ranking scale. If the order is related to given criteria, it may be called a rating scale. Different performances are often assigned numerical marks and the marks added. The numbers may then be treated like measurements of length or weight, as if, for example, 50 marks represented twice the performance of 25 marks. Mark scales can be misleading in this respect.

Norm-referencing

An assessment system in which pupils are placed in rank order and pre-determined proportions are placed in the various grades. It implies that the grade given to a particular pupil depends upon a comparison between that pupil's performance and those of all the other pupils in the group, rather than upon the absolute quality of the performance.

Criterion-referencing

As defined for this report, an assessment system in which an award or grade is made on the basis of the quality of the performance of a pupil irrespective of the performance of other pupils: this implies that teachers and pupils be given clear descriptions of the performances being sought. This is a broader and less exacting definition than that used by some authors.

Aggregation

Combining a pupil's marks or grades on a set of different assessment items to give a single mark or grade. This may be done in different ways: for example, two numerical marks may be added together, or two grades may be combined according to rules (eg. AA, AB or BA all give a complete A).

Level

In this report, level indicates achievement of a particular set of performance criteria which have been defined within a "profile component" (see section VI), or cluster of attainment targets. In general, a higher level will imply achievement of greater sophistication or range than that for a lower level.

Grades

Different qualities of response to a common assessment demand or task. "Pass/fail" examinations are usually referred to as ungraded. Criterion-referenced assessments are "pass/fail" insofar as pupils' performances are judged with respect to unitary criteria. The equivalents to grades in a criterion-referenced system are results summed over the attainment target criteria, so that success is judged relative to those attainment targets which have been met. In this report we distinguish between the quality of response to the same attainment targets (grades), and the degrees of success in reaching a collection of attainment targets (levels).

Moderation

The process of checking comparability of different assessors' judgements of different groups of pupils.

The term can also refer to a procedure for determining a single set of marks or ratings when two sets are available for the same pupils on the same profile components, eg. one set from a teacher's course assessments and one from a national test. One set (usually the teachers') may be adjusted to give the same mean and spread as the other. This is a statistical operation usually called scaling. There are other forms of statistical moderation: for example, when the "results" are expressed as grouped percentage distributions of pupils, adjustments can involve deciding which borderline pupils should change their group status. This can be a more sensitive procedure than scaling.

Calibration

Where procedures for equating results of the assessments are statistical, the procedure is sometimes called calibration, to distinguish from procedures in which the judgement of assessors is involved (moderation). This scaling is a form of calibration. Procedures using both statistical data and judgements are sometimes called "intelligent statistical moderation".

Skill

Ability to perform, usually acquired by training: can be displayed in many types of performance, eg. manual, craft, design, communication, observation, measurement.

Concept

A generalised idea or notion; identifying something as an example of a particular concept enables one to infer how to treat it.

Knowledge

That which is known in an ordered way and can be used; in this sense different from just a collection of unrelated information.

Understanding

The ability to select and use knowledge and skills over a variety of contexts to meet a variety of demands.



King's College London (KQC)

UNIVERSITY OF LONDON

CHELSEA CAMPUS

552 KING'S ROAD, LONDON SW10 0UA TELEPHONE: 01-351 2488

CENTRE FOR EDUCATIONAL STUDIES

HEAD OF DEPARTMENT
Professor P. J. Black, OBE, KSG

24 December 1987

Rt Hon Kenneth Baker MP
Secretary of State for Education and Science
Department of Education and Science
Elizabeth House
York Road
London SE1 7PH

Dear Secretary of State

TASK GROUP ON ASSESSMENT AND TESTING

I present herewith the main report of the Task Group which you set up last July to advise on assessment and testing within the national curriculum.

Because of the short time available we have not been able to give adequate attention to one part of our remit: we are not yet in a position to make recommendations about the various services and arrangements required to support the system that we propose. We discuss some of the issues involved in the enclosed report, but we would like to give more considered advice about these in about two months time. We might also wish at that stage to give fuller elaboration on other issues, as envisaged in your letter of guidance sent to us last October.

I am happy to assure you that this report has the support of all members of our group. I am grateful to many members of your Department for the help and support that they have given during our work.

Yours sincerely

Professor P.J. Black
Chairman
Task Group on Assessment and Testing

REPORT OF THE TASK GROUP ON ASSESSMENT AND TESTING

INTRODUCTION

I OUR STARTING POINT

1. The Task Group has been asked to advise on the practical considerations governing assessment within the national curriculum. We start from the following propositions.
2. A school can function effectively only if it has adopted:
 - clear aims and objectives;
 - ways of gauging the achievement of these;
 - comprehensible language for communicating the extent of those achievements to pupils, their parents and teachers, and to the wider community, so that everyone involved can take informed decisions about future action.
3. Promoting children's learning is a principal aim of schools. Assessment lies at the heart of this process. It can provide a framework in which educational objectives may be set, and pupils' progress charted and expressed. It can yield a basis for planning the next educational steps in response to children's needs. By facilitating dialogue between teachers, it can enhance professional skills and help the school as a whole to strengthen learning across the curriculum and throughout its age range.
4. The assessment process itself should not determine what is to be taught and learned. It should be the servant, not the master, of the curriculum. Yet it should not simply be a bolt-on addition at the end. Rather, it should be an integral part of the educational process, continually providing both "feedback" and "feedforward". It therefore needs to be incorporated systematically into

teaching strategies and practices at all levels. Since the results of assessment can serve a number of different purposes, these purposes have to be kept in mind when the arrangements for assessment are designed.

5. Any system of assessment should satisfy certain general criteria. For the purpose of national assessment we give priority to the following four criteria:

- the assessment results should give direct information about pupils' achievement in relation to objectives: they should be **criterion-referenced**;
- the results should provide a basis for decisions about pupils' further learning needs: they should be **formative**;
- the scales or grades should be capable of comparison across classes and schools, if teachers, pupils and parents are to share a common language and common standards: so the assessments should be calibrated or **moderated**;
- the ways in which criteria and scales are set up and used should relate to expected routes of educational development, giving some continuity to a pupil's assessment at different ages: the assessments should relate to **progression**.

6. Our recommendations have been composed in the light of these. Unless the criteria are met, the potential value of national assessment in assisting learning and supporting the professional development of teachers is unlikely to be realised.

II PRESENT PRACTICE

Assessment in schools

7. Generally speaking, teachers and pupils are well aware of the importance of assessment. A great deal of effective assessment is already carried out, often in an informal manner within the individual classroom. Many schools have established common systems for use throughout the school. In primary schools

these systems typically require agreement and joint action by all the teachers in the school, while in secondary schools collaboration is focussed within subject departments to establish common methods and criteria. Where this happens the results become one of the main inputs in a review of the work of a school or a department. Elsewhere, common guidelines may be used throughout a school. A shift from norm-referencing to criterion-referencing improves the value of this. Systematic records of pupils' progress can then be helpful in the transition between year groups and between schools. Records become the basis for discussion with pupils, and with their parents. For all of these uses assessment which provides details of achievement in different aspects of any subject is more valuable than a single overall grade.

8. Teachers' own use and interpretation of assessment and testing data is, however, limited by the range of instruments available, by the lack of time and resources required to develop new instruments, and by the lack of external reference and a basis for comparison. External agencies involved with teachers in developing the provision and use of new assessment and testing procedures - for example through the Assessment of Performance Unit, the various new Graded Assessment schemes, or in new GCSE work - have found that both primary and secondary teachers welcome the professional development flowing from better assessment methods and exchange of information about expected outcomes. Experience, particularly but not exclusively in mathematics, suggests that paying greater attention to assessment, with improved means for assessing progress in relation to curriculum targets, helps teachers to learn important lessons about their pupils. This often reveals that the needs of the less able, or the competence of the most able, have hitherto been under-estimated.

External initiatives

9. Many helpful developments in assessment have arisen from initiatives outside the individual school, including those of local education authorities, the Department of Education and Science, the Secondary Examinations Council and the new examining groups. These show that initiatives at a national level which lead to production and development of assessment instruments, and marking and moderation procedures, can be of great value within schools if carefully considered and sensitively deployed.

10. Large scale use of external standardised tests is not a new or unusual practice. In the primary phase, almost all children take such tests at some time (see Appendix C). Our evidence has shown however that these externally-provided measures are not usually related, in either construction or subsequent use, to the normal work of teachers. Classroom assessments are hardly affected by them. In addition, whilst some of the more recent tests reflect good assessment practice, many older but still widely used tests are narrow in scope and employ a limited set of techniques. Although external tests have been of some value in screening to identify children with special learning needs, they have rarely been used to assist normal teaching work. The resources devoted to such testing have been scarcely justified, and the more limited types of test have themselves given testing in general a reputation for narrowness and irrelevance.

Local authority systems

11. Many local authorities have developed guidelines, and even common instruments and systems, for assessment in their schools. Some require use of standardised tests by all schools and treat the results as performance indicators. The difficulty here is commonly the lack of relationship between these tests and the learning aims actually pursued, so that teachers often regard the tests as time-wasting, or even oppressive interference. Other authorities provide only very general guidelines, believing that individual teachers must take full responsibility for such work. The difficulty here is that little external help can be provided. There is therefore no shared language and no fostering of common perceptions and criteria to help teachers and schools to improve their work, or to help parents and others to understand the performance of a school. Authorities rarely seem to have achieved an effectively integrated system in which centrally provided assessment resources give useful overall information, and also help by indicating appropriate methods and providing calibration, which teachers can make use of in improving their own work in their own ways. Several, however, are seeking to develop systems with these characteristics.

Other countries

12. Many other countries have a national curriculum or uniform curricular guidelines for each region. Many have examinations at the end of compulsory

schooling, and for selective access to further education, but only a few have national assessment schemes for younger ages. In most cases, assessments at these younger ages are left to teachers. In almost all European countries the regular assessment by teachers, using mainly written work which they set and mark, is designed to indicate pupils' progress in relation to a national or regional curriculum. Results of internal assessments are sometimes communicated to parents, according to national or local law, in terms of grades on a specified scale (1 to 5 in Sweden with 5 being high, 1 to 6 in West Germany with 6 being low). In West Germany however these grades are only loosely defined and there is no procedure for moderating assessment or for calibration between schools. In Sweden standardised achievement tests are used for calibration; these are compulsory beyond 14, voluntary - but widely used - from about age 9 to 14. Publication of test results for use outside the school is virtually non-existent, except in some States in the USA. There results have been published with adjustments for socio-economic background effects: but while this can give a fairer reflection of the achievements of a school, it has made the information too complicated for the general public to understand. No country appears to have a national assessment system which is well developed in relation to formative purposes and to a framework of progression. It is worth noting also that assessment instruments developed by the APU in this country for monitoring purposes are internationally recognised for being both innovative and of the highest quality.

III CONCERNS TO BE ADDRESSED

13. Whilst present practices, both in this country and in others, give many examples of positive uses of assessment, no system has yet been constructed that meets all the criteria of progression, moderation, formative and criterion-referenced assessment set out in paragraph 5 above. Our task has therefore been to seek to devise such a system afresh. We believe that the model of assessment put forward in this report builds on some existing good practice and represents an advance on assessment practices in other countries. But we could not approach this task without also recognising that many are deeply opposed to any system of national assessment and testing. They fear that it may do serious damage and they see no possible advantages. In forming our recommendations, we have tried wherever possible to address these fears, in ways we summarise below.

Effects on pupils

14. Some of the fears relate to **damage to pupils**. Yet when assessment and testing are carefully aligned to the curriculum, as in Graded Assessment Schemes, one of the outstanding benefits that teachers report is the enhanced motivation of pupils. Many pupils seem to need short-term objectives which are clear and which they perceive to be attainable. The possibility of achieving these builds enthusiasm for progress in learning. Of course pupils do not always succeed, but poor performance against a target should not be seen as a prediction of personal inadequacy; it should usually be regarded as an indication of needs. Such needs must be met, perhaps first by the use of more detailed diagnostic assessments conducted usually by the school to identify specific difficulties, and then by further teaching help. It would be absurd if diagnosis of difficulty - and recognition by a pupil and the parents and teachers of the need for help - were to be avoided because of fear of failure. It would equally be wrong if such evidence about an individual were to be reported to others beyond the parents, except in confidence to those in a position to provide help. Our proposals reflect these considerations and, we believe, can positively assist children's development.

Relations between parents and schools

15. Other fears relate to possible **damage to the relationships between parents and schools**. But both parents and pupils need more intelligible information than is provided at present, particularly in the secondary phase where a parent may receive reports from several teachers, each using different scales and language to record progress. Parents are interested above all in where their child stands in terms of the general progress expected at that age. At present the usual comparison, if any, is with the class or year group within the school. A national curriculum leading to national criteria and calibration could change this, and could help schools to give firmer information and where appropriate, reassurance to parents. Similarly, employers and further education institutions, who often have to select in advance of GCSE results, cannot readily relate the reports of different schools because they cannot compare the standards applied by those schools and the different scales and levels of detail used in reports. Our proposals aim to promote and support an informed relationship between schools, parents and others.

Effects on teachers' work

16. A more widely-voiced fear is that external tests will impose arbitrary **restrictions on teachers' own work**, and so limit and devalue their professional role. Our first response to this has been to recommend procedures in which:

- teachers' assessments over time and in normal learning contexts play an important part;
- externally-provided methods and procedures are broad in scope and related to the curriculum attainment targets which all will share;
- these assessment methods may often be incorporated into normal classroom activities;
- the administration, marking and moderation procedures rely on the professional skills and mutual support of teachers, giving them both key responsibilities and communal safeguards against idiosyncrasy.

17. But we also recognise that many teachers, at all levels, will need help to incorporate any new and systematic assessment procedures into their normal classroom work. Experiences both in assessing GCSE subject course-work of pupils at ages 15 and 16 and in in-service training for primary teachers working with 7 year olds on project tasks have shown that teachers need time and support to help them adapt. They have also shown that teachers welcome the professional growth they gain from such work if they can come to share the assessment aims and translate the methods into their own style of work. These considerations indicate that provision of carefully planned and pilot-tested assessment instruments, and of in-service training directed to initiating and supporting new types of classroom work, will be essential.

Publication

18. Finally, there is a fear that results will be published in **league tables of scores**, leading to ill-informed and unfair comparisons between schools. We believe that most teachers and schools would not object to assessment results being reported to those who know the school and can interpret them in the light

of a broader picture of its work and circumstances. They would object, however, to the publication of partial information which is not set in that context and is therefore potentially misleading - particularly where significant decisions are then based upon that information. It needs to be recognised that many of the outcomes of a school's work are influenced by external circumstances, notably the social and economic background of pupils' families, over which the school has no control. These factors do not of themselves affect the validity of a criterion-referenced attainment target. But judgements about the quality of a school should not be confined to the extent to which the targets are actually reached. They should also take into account the educational value added - that is, the progress it might have been reasonable to expect a school in such circumstances to secure among its pupils, bearing in mind the pattern of attainment at intake and variations in children's rates of development. We recognise that these factors are difficult to measure and interpret; this strengthens the need for information about a school's environment. We therefore consider that the publication of assessment results for any school should take place if, and only if, this is done in the context of reports about that school as a whole, so that it can be fair to that school's work and take account so far as possible of socio-economic and other influences.

Our proposals

19. No new scheme could immediately set all such fears entirely at rest. We emphasise, however, that many of them arise from assumptions, based on experience of previous schemes and practices, which do not apply to our proposals. Our proposals are both evolutionary and radical. They are evolutionary in that they build on examples of good practice that we have studied, some of them of recent origin. They are radical in that they are developed and will be adopted on a national scale, and because our remit can be fulfilled in a positively helpful way only by some quite new departures. A more detailed discussion of this important issue is given in Section XXII.

IV THE STRUCTURE OF THE REPORT

20. With those considerations in mind, the following sections of this report set out the arguments and our recommendations for a national system of assessment and testing. Section V starts by discussing the purposes of the

assessment system. This introduces a sequence of sections VI - X exploring the ways in which any particular performance component, representing a linked cluster of attainment targets within a subject, should be assessed and moderated. The next sequence XI - XV considers how these components may be combined to form overall assessments for pupils at and across various ages, how they should be reported, and the implications for primary and secondary schools and special educational needs. This is followed by sections XVI and XVII which set out various consequences for the tasks of the national curriculum subject working groups and for phasing the introduction of national assessment arrangements. After this, we turn in sections XVIII - XXI to implications for the setting up and support of the national system. Finally in sections XXII-XXIII we discuss some general conclusions and summarise our recommendations.

21. The report is accompanied by appendices. Appendices A, B and C respectively give the Group's terms of reference and membership, the Secretary of State's letter of guidance, and a list of meetings held. Subsequent appendices expand on issues which, for brevity's sake, have to be presented in the main report without full supporting argument or illustration.

22. We draw attention also to the need for clear terminology. The preface to this report gives the necessary specifications and these have been followed throughout; in many cases common usage is more flexible, but for our purposes this could have given rise to misleading ambiguity.

PURPOSES AND PRINCIPLES

V PURPOSES

23. The terms of reference and the letter of guidance from the Secretary of State require that information derived from assessments (including tests) shall be capable of serving several purposes:

- formative, so that the positive achievements of a pupil may be recognised and discussed and the appropriate next steps may be planned;

- diagnostic, through which learning difficulties may be scrutinised and classified so that appropriate remedial help and guidance can be provided;
- summative, for the recording of the overall achievement of a pupil in a systematic way;
- evaluative, by means of which some aspects of the work of a school, an LEA or other discrete part of the educational service can be assessed and/or reported upon.

24. Our attention has also been drawn to the effects all forms of assessment have on teaching and learning. We are required to propose a national assessment system which enhances teaching and learning without any increase in "the calls on teachers' and pupils' time for activities which do not directly promote learning". We judge that it would be impossible to keep the burdens on teachers and pupils within reasonable bounds if different batteries of assessments and tests had to be created to serve each separate purpose. We must therefore reflect the priority given in our brief to the need "to show what a pupil has learned and mastered".

Serving several purposes

25. Some purposes may, however, be served by combining in various ways the findings of assessments designed primarily for a different purpose. It is possible to build up a comprehensive picture of the overall achievements of a pupil by aggregating, in a structured way, the separate results of a set of assessments designed to serve formative purposes. However, if assessments were designed only for summative purposes, then formative information could not be obtained, since the summative assessments occur at the end of a phase of learning and make no attempt at throwing light on the educational history of the pupil. It is realistic to envisage, for the purpose of evaluation, ways of aggregating the information on individual pupils into accounts of the success of a school, or LEA, in facilitating the learning and achievements of those for whom it is responsible; again, the reverse process is an impossibility.

Formative and summative

26. We judge therefore that an assessment system designed for formative purposes can meet all the needs of national assessment at ages before 16. At age 16 the focus shifts from formative to summative. The GCSE is a terminal examination in the sense that it comes at the end of the compulsory period of education and thus cannot avoid summative judgements. Similar considerations would apply to other forms of assessment at that stage. Summative judgements or aggregations may be made at ages other than 16, but only at 16 does it seem appropriate for assessment components to be designed specifically for summative purposes.

Formative and diagnostic

27. We do not see the boundary between the formative and diagnostic purposes as being sharp or clear. If an assessment designed in formative terms is well matched to the pupil, it is likely to provide some information which will help in the diagnosis of strengths and weaknesses. Some assessments of this kind will be of value as indicators of learning problems which require further investigation. Further diagnostic investigations would, however, often involve the use of highly specialised tests of relevance to only a limited set of circumstances and pupils. It follows that the results of such tests should not be included in aggregated results for groups of pupils. Furthermore, a detailed diagnostic report on every pupil would lead to an excess of information and hence be counter-productive. The formative information about pupils will be passed on to their next teacher; if diagnostic reports are confined to problems outside the normal range, these will be highlighted and made more salient to the receiving teacher.

We recommend that the basis of the national assessment system be essentially formative, but designed also to indicate where there is need for more detailed diagnostic assessment. At age 16, however, it should incorporate assessment with summative functions.

Overall results for a pupil

28. The complete range of formative and diagnostic purposes will be served only if the results for each pupil are aggregated in various ways: for example, across cognate elements in part of a subject or field of study or across the whole of that subject or field. There may also be value in comparisons which place the performances of the pupil in the contexts of his or her earlier achievements and the achievements of the class as a whole. Out of these aggregations and comparisons can come judgements to be presented to the pupil and/or parent so that they are involved in making decisions about the pupil's future educational programme. Such judgements raise questions of the confidentiality of information about individuals. We see confidentiality as crucial if assessment is to serve a positive role.

We recommend that all assessment information about an individual be treated as confidential and thus confined to those who need to know in order to help that pupil.

Evaluation

29. The other purposes listed in paragraph 23 can be fulfilled by shifting the focus from individual to class, or to a group of classes in a school, or to the school, or to a group of schools within an LEA. The same processing procedures of aggregation and comparison, together with an analysis of the spread of the results around the average, can be applied to yield information of value to a range of audiences. National assessment designed to have formative/diagnostic functions and to serve as a foundation for all the purposes will give results which have a common basis in the progress of the pupils in relation to national targets.

We recommend however that for summative and evaluative purposes results should be aggregated across classes or schools so that no individual performances can be separated out.

Assessment of attitudes

30. We draw attention to the fact that Records of Achievement serve purposes which overlap but are not identical with those discussed above. In our judgement it is appropriate for such records to include information about the attitudes of pupils, since statements of this kind will have been discussed with the pupil and inserted by agreement into a comprehensive record which is the property of, and confidential to, the pupil. But because of their wider role, the assessments associated with the attainment targets of the national curriculum need not, and should not, be subject to quite such close confidentiality as may be appropriate for Records of Achievement as a whole or for the personal components which are included in such records. We recognise the importance of educational aims which relate to the development of personal attitudes, and so accept that these might be considered by subject working groups in framing attainment targets. We judge it to be wrong, however, to require any part of the national system of assessment, including reporting, to be concerned with this part of the educational programme.

We therefore recommend that assessment of attitudes should not form a prescribed part of the national assessment system.

VI PROFILE COMPONENTS IN THE ASSESSMENT FRAMEWORK

31. The basic unit of information in any assessment system is the response of an individual pupil to a particular demand or task: for example, a written response to a test item, an observed action in a practical task, or a sequence of actions assessed by a teacher's rating. To base an interpretation upon only one unit of information is neither sufficient nor desirable. Any particular unit of information will depend heavily on the context in which the information is gathered. Before any such response can be interpreted with any certainty, it must be compared with the responses of the same pupil to other tasks and perhaps with the responses of other pupils to the same tasks. The standard method of making these comparisons is by aggregating the responses.

Feeding forward

32. The formative and summative purposes of assessment, concerned with the immediate needs and achievements of the pupil, necessarily involve aggregations

of the responses to a range of tasks by individual pupils. As we have already said, the formative purpose is fundamental to the national assessment system. In order to serve that purpose effectively, the results must both feed back to the pupil and feed forward to the next teacher or institution. The emphasis on the formative function serves to underline the "feed forward" aspect. This is enhanced by the choice of reporting ages, which carries with it the implications that the teachers responsible for making the assessments will not be those responsible for acting upon them.

33. This feed forward role is diminished if the assessment of a pupil's progress in a subject is designed to produce only a single overall score. For example, in developing graded assessments in science, the London Group chose to assess and report separately on three components - skills (mainly practical), knowledge and understanding, and exploration (project) work. The results in these three components form a profile which represents the pupil's attainment in science. Similarly, many schemes for English recognise the need to attend separately to writing, oracy, reading comprehension, and listening: again, results on these four would give a profile. Information of value to pupils, parents and others would be lost if (say) a teacher were to add the results of writing to those of oracy, and to ignore those separate results when discussing pupils' future work. If subject attainments are presented as a profile of separately determined scores, then receiving teachers will be better able to use the information to shape their teaching strategies.

We recommend that, to realise the formative purpose of the national assessment system, pupil results in a subject should be presented as an attainment profile of the kind discussed above.

Profile components

34. The profile will thus be composed of a set of profile components. These components should be broadly based so as to be informative to the teachers of a range of subjects being studied by the pupil. In addition, each reporting age may herald the introduction of new subject areas. For these in particular, teachers will want a report on subjects already studied which goes beyond subject-specific knowledge and skills to the general skills, knowledge and

understandings upon which they can build. Including in a subject profile those components which have applicability beyond that subject will serve to emphasise the underlying unity of a subject-based curriculum.

35. Defining the profile components will be the responsibility of the relevant subject working groups. We envisage that each component will comprise a cluster of attainment targets which have some homogeneity in relation to the skills, knowledge and understanding which the subject promotes. But their internal consistency should not be the only determining factor. To demand high correlation between scores on all of the attainment targets within each profile component might lead either to an unmanageable number of profile components, or (worse) to the omission on statistical grounds of educationally important targets. There is a balance to be struck. On the one hand, a large number of components would daunt teachers and confuse parents. On the other, a small number would mean that each component might be too heterogeneous to be meaningful.

We recommend that an individual subject should report a small number (preferably no more than four and never more than six) profile components reflecting the variety of knowledge, skills and understanding to which the subject gives rise. Wherever possible, one or more components should have more general application across the curriculum: for these a single common specification should be adopted in each of the subjects concerned.

36. Each profile component ought to be assessed separately and separately moderated. This has implications for the amount of assessment and moderation overall. We discuss in section XIII below the danger of overburdening primary school teachers in particular.

Feeding back

37. While the above argument for profile components rests on the feed forward aspect, profiles also provide feedback. The successes of the graded assessment movement point to the educational value of detailed feedback to pupils. Attainment targets, shared by teachers and pupils, appropriate to the current

stage of the pupil's achievement and pointing the way forward to further progress, appear to play an important part in raising standards of achievement. Thus formative assessment must be in terms to which pupils can respond and which clearly indicate what has been and what remains to be achieved. The profile components should give meaningful information for this purpose and help to convey and draw upon the diversity of performance that a pupil may exhibit in any one subject.

Aggregation for evaluation

38. These arrangements should serve the evaluative purposes of national assessment equally well. Aggregating pupil performances to the level of the class, the school, the LEA, or the age group should help to monitor the delivery of the national curriculum. These aggregates can be based on any unit of analysis. Aggregates of performances in test items, standard assessments, attainment targets, profile components, whole subjects or larger curriculum segments all have their evaluative uses.

39. To sum up so far we envisage that the outcomes of the national assessment system will be presented as the attainments in a set of profile components some cross-curricular, some subject specific. The profile will become the main vehicle of communication in helping both pupils and teachers to evaluate their work.

VII ASSESSMENT AND TESTING METHODS

40. The process of education has the aim of changing people and their behaviour by imparting skills, increasing knowledge and understanding, and developing competences. This is a dynamic process in which individuals are constantly developing. It is important for teachers to be aware of the levels reached by pupils, and their rate of development, in order to facilitate learning.

Scope of assessment

41. As a natural part of teaching, therefore, teachers are constantly assessing pupils to determine their progress and to plan the next stage of their learning.

In the widest sense, such assessment involves a continuous comprehensive examination of all aspects of the pupil's learning, drawing on a wide variety of evidence from many sources to arrive at a general picture. This process is not sharply focussed, yields a diversity of information and tends to be only loosely related to purpose. Consequently it may not be clear how it relates to particular decisions about a pupil's educational programme.

42. In a narrower sense, the term 'assessment' is used to refer to an individual component of the total assessment process or to a particular method of assessment. Hence, it encompasses all procedures used to make an estimate or appraisal of an individual's achievement. Which of the many methods of assessment may be appropriate in particular circumstances will depend on the purpose of the assessment and the decisions to be taken in the light of the results obtained.

43. It will be clear that there are many sources of information for assessment, each of which is more or less suited to different purposes. Some of these sources are:

- general impressions;
- marking course work;
- marking assignments;
- pupils' self-assessment;
- rating scales;
- checklists;
- practical tests;
- written tests.

These have differing levels of formality and differing levels of standardisation. Some are appropriate only for making comparisons within a particular class or school (impressions, course work); others are more generally applicable. Some assessment methods provide a richness of detail and a validity in context which cannot be obtained in more formal ways. However, these do not usually provide a manageable means of contrasting the performance of pupils who are being assessed by different teachers, whether in the same school or in different schools.

Standardisation

44. In order to secure comparability, an assessment component needs to be conducted according to formal procedures related to the particular aspects of an individual's knowledge, skills and competences being examined. A standardised assessment of this sort would be designed to elicit a particular type of response in a standardised manner, and to rate this, using standard methods, so that the pupil assessed would be assigned a score, grade or category. Standardisation allows the assessment component to be used widely by many different teachers who, if they are appropriately trained in its use, can thereby produce results which are on a common scale. But it is usually time-consuming. So if the calls on teaching time are to be contained, standard procedures can be used by teachers for only a limited proportion of a pupil's work, and the need to standardise may narrow the range of attainments being assessed.

Testing

45. At this point we have a difficulty with terminology. Our terms of reference use the words **assessment** and **testing**. The former, as explained above, has a broad inclusive meaning. The meaning given to **testing** is often too narrow. Much of the discussion about testing seems to be based on a particular image. This image is frequently of an externally-prescribed paper and pencil test, formal and unimaginative in scope, to be attempted in a set time on a formal occasion and marked according to set rules. The system that we recommend must include the use of tests, but far more broadly conceived. The aim of the system must be to provide standardised, ie nationally comparable, assessment results; but the range and scope of standardised assessments that can be used to perform this role is far wider than the term "test" is usually taken to imply. We emphasise this point to draw attention to the fact that our use of the word **test** will have this broader meaning. Ideally it might be better expressed by the phrase "**standard assessment task**". Both terms will be used interchangeably in this report, to specify externally provided tasks and procedures designed to produce performance data on a national scale.

46. As we have said, teachers carry out assessment using a wide range of methods, including tests. The national system that we propose will use, in

combination, teachers' own assessment results together with the results they obtain with tests, ie with standard assessment tasks.

Variety of tasks

47. Standard assessments need not only be in written form. Indeed, the wide variety of possibilities can be explored by analysing any one task in terms of three aspects or modes, which can be defined as follows:

- the **presentation** mode - the method of delivery of the questions (oral, written, pictorial, video, computer, practical demonstration);
- the **operation** mode - the expected method of working (mental only, written, practical, oral);
- the **response** mode - pupils may answer in various ways (eg choosing one option in a multiple-choice question, writing a short prescribed response, open-ended writing, oral, practical procedure observed, practical outcome or product, computer input).

48. The art of constructing good assessment tasks is to exploit a wide range (far wider than those normally envisaged for tests) of modes of presentation, operation and response, and their numerous combinations, in order to widen the range of pupils' abilities that they reflect and so to enhance educational validity. Examples of the wide range of possibilities are set out in Appendix E. These examples, which are all in current use or under development should serve to illustrate our view that tests, understood as standard assessment tasks, can be devised to achieve educational validity in relation to a wide range of learning aims.

49. One aspect of this expanded range of possibilities is that the discontinuity between teachers' own assessment of normal classroom work and the use by them of externally provided tests need not be a sharp one. For example, an open-ended task - presented in a prescribed way in a practical mode, requiring practical and oral work and assessed in a standard way through oral and written response by the pupil - could well be so deployed by a teacher that

pupils would not necessarily be aware of any departure from normal classroom work (see Appendices D and E for examples).

Flexibility

50. Provided the requirements of national comparability are met, the system we propose allows a degree of flexibility at the school level. For example where provided externally, the particular test topics or themes may be a prescribed set, or an agency may provide a range from which schools can make their own choice. The form in which a task is administered may be prescribed, but it may not have to be done at fixed times - particularly for tests of a practical project nature which may extend over several class periods. If such flexibility is to be accommodated - which we think very desirable - then marking or grading procedures will have to be carefully standardised, and there will also have to be some means of moderation between teachers and schools (see section X).

We recommend that the national system should employ tests for which a wide range of modes of presentation, operation and response should be used so that each may be valid in relation to the attainment targets assessed. These particular tests should be called "standard assessment tasks" and they should be so designed that flexibility of form and use is allowed wherever this can be consistent with national comparability of results.

VIII BIAS IN ASSESSMENTS

Gender and ethnic bias

51. As with most assessments, the results of the standard assessments are likely to show differences between groups. It is possible that there will be sex differences in some attainment areas; for example, on average girls score more highly than boys in certain writing tasks. Generally, though, the position is not simple - in mathematics boys do better in some topic areas and girls in others. There is also some evidence that different types of test favour one or other group (eg boys may do better in multiple-choice tests). In Appendix F we

reproduce a paper from the Equal Opportunities Commission which discusses such problems in detail. Similarly, problems of bias may arise in relation to differences between the various ethnic groups.

52. Two procedures should, in our view, be incorporated into the processes of development and analysis of standard assessments. Firstly, the agencies responsible for developing the standard assessments should incorporate formal reviews of proposed material for any evident sex or race stereotyping, and ensure a proper balance. This would normally be carried out in collaboration with panels of teachers. Secondly, the assessment devices should be statistically examined for evidence of item bias (a form of invalidity) and as far as possible the sources of such bias should be eliminated. Established statistical procedures exist for the detection of item bias.

We recommend that assessment tasks be reviewed regularly for evidence of bias, particularly in respect of gender and race.

Language difficulties

53. There may be difficulties presented by those whose first language is not English. Head teachers might exempt such children from tests where this problem is so severe as to render the assessment unworkable. Yet it should be recognised that to record a low level of performance for this reason would be no reflection on a pupil's general ability but merely an indication that the pupil needed special help in English language skills. Assessment in other skills and understanding, particularly at age 7, should, wherever practicable and necessary, be conducted in the pupil's first language.

IX ENSURING CONFIDENCE IN NATIONAL ASSESSMENT

54. The national assessment system must be designed to yield reports of pupil performance in which a wide variety of users can have confidence. This confidence arises from the reliability of the observations made and the validity of the interpretations placed on them. Both reliability and validity contribute to a statistical estimate of confidence. In both cases confidence can be increased by using more, and more varied, types of test or task. Repetition of assessments is, however, a potential burden on teachers and pupils alike and can

distract from teaching and learning. The national assessment system must therefore minimise the amount of information to be collected while maximising confidence in its interpretation.

55. Drawing the conclusion that a pupil has achieved the objectives of the national curriculum involves a chain with five main links; each carries the possibility of error.

The national curriculum must be clearly communicated

56. The first risk to confidence arises from lack of clarity in the definition of what has to be learned and assessed. In the past what is to be assessed has often been the only clear expression of what is to be taught and this has often led to a narrowing of the curriculum. This tendency can be reduced if the process can start from agreement about what has to be learned in terms of attainment targets. Illustration of what these mean, and imply, can then be communicated most clearly by means of specimen tasks with sample responses.

We recommend that attainment targets be exemplified as far as possible using specimen tasks. Such tasks can then assist in the communication of these targets.

57. Clarity of definition should not be confused with reduction to a few key elements. The curriculum is inherently complex and the subject working groups should be encouraged to mirror that complexity. We assume that there will be more attainment targets to be met at any reporting age than can reasonably be assessed on a single occasion for national assessment purposes. This problem affects both the choice of methods and the number of sampling occasions. These two issues are discussed separately in paragraphs 58 to 60 below.

National tests must use a variety of methods

58. Secondly, because targets may have to be sampled, this carries the danger of narrowing the scope of assessment. This danger is enhanced by the fact that some kinds of tasks are easier and cheaper to set and mark than others. Use of only one kind of assessment instrument in the interests of economy can lead to undue emphasis on some targets at the expense of others, threatening distortion

of curriculum coverage. A broad range of assessment instruments sampling a broad range of attainment targets will further discourage the narrowing tendency to "teach to the test".

59. We envisage that tests in a particular subject might cover up to half a dozen profile components. The information gained in what must be a limited time can be increased if items cover (say) application of knowledge and are designed to give information both about the knowledge and about the skills of application. Written test responses might be supplemented by teacher observations of skills made in a systematic way in the same or similar externally defined contexts. Together these can provide a sample covering the entire syllabus. The assessment should include items and activities in which pupils have to supply extensive ideas and information in their responses; error analysis of such data can be very revealing of pupils' learning difficulties.

We recommend that a mixture of standardised assessment instruments including tests, practical tasks and observations be used in the national assessment system in order to minimise curriculum distortion.

Observed responses must be shown to be typical

60. The third factor is that pupils' responses to tasks are influenced not only by their understanding of what the task involves but also by how and when the task is presented. Pupils perform differently on different days and in different circumstances. Any single response is open to these risks. They are minimised in teachers' ratings of longer term performance because those are based on many occasions and many different types of task. Teachers' ratings have the further advantage of flexibility. They can be focussed on the targets of the national curriculum without unduly restricting the school's choice of ways of meeting these targets. That freedom of professional decision is an important aspect of the national curriculum. Externally prescribed tests or tasks cannot sample responses over many occasions because this would require undue interference with teaching and learning.

We recommend that teachers' ratings of pupil performance should be used as a fundamental element of the national assessment system. Just as with the national tests or tasks, teachers' own ratings should be derived from a variety of methods of evoking and assessing pupils' responses.

Performance must be interpreted in terms of achievement

61. Fourthly, the effects of context described above are compounded, and thus likely to be confused, with more fundamental differences in pupil performance. Targets for any reporting age will be achieved by different pupils in different sequences. If reliable results were to be required about every single attainment target, an excessive amount of assessment would be needed in order to disentangle the effects of context from more fundamental individual differences in the sequence of achieving targets. But profile components based on aggregations of target achievements can be reported with acceptable degrees of confidence. Thus the case for specifying only a small number of profile components in order to support clear communication is reinforced by the argument presented here that greater confidence follows from the aggregation of targets into profile components.

We recommend that when the subject working groups provide guidance on the aggregation of targets into a small number of profile components, they should have regard to the need for each component to lead to a report in which reasonable confidence is possible.

Interpretations of achievement must be related to national standards

62. Finally, national standards are directly conveyed in the mark schemes of standardised tests. If teachers' rating scales are to communicate national standards clearly, they should not only report on a small number of profile components but also use a short scale for each component so that each scale point can be tied to a description of a significant level of attainment. Teacher ratings are, however, open to misinterpretation, since they are strongly influenced by local expectations of achievement. On the other hand, teachers'

expectations have an important role in providing feedback to help to ensure that national standards, based on subject-specific requirements, are realistic. Moderation is the key to effective national standardisation. We discuss what moderation involves in the next section.

We recommend that teachers' ratings be moderated in such a way as to convey and to inform national standards.

63. To achieve confidence in the face of these risks of error is not such a demanding requirement as at first appears. A combination of teachers' ratings and standardised assessment tasks can engender confidence in the interpretation of achievement. The combination will facilitate fair sampling of the national curriculum, coverage of different circumstances of assessment on different occasions, and a range of evidence as the basis of report. Individually the instruments are open to misinterpretation but in combination they minimise the risks.

We therefore recommend that the national assessment system is based on a combination of moderated teachers' ratings and standardised assessment tasks.

X MODERATION

Purpose of moderation

64. The process of bringing individual judgements into line with general standards is known as moderation and is an essential part of an assessment system.

65. We are all, as individuals, persuaded that those things which occur frequently in our experience are 'normal'. Schools' catchment areas are not representative of the national population but they condition teachers' expectations of what is a 'normal' average and spread of achievement. These expectations are powerfully reinforced by successive year groups of pupils. In the absence of equally powerful external evidence, teachers' expectations become the teachers' standards.

66. Teachers' expectations of individual pupils in the class create further problems, loosely referred to as the 'halo' effect. In the absence of a close definition of what to look for and how to observe it, we look for confirmation of our expectations. Research evidence shows that teachers' rank orders of pupils' performance may vary systematically from rank orders provided by test users.

67. All three statistical features of a set of school-based assessments - the mean, the spread and the rank order - are affected by fallibility in judgement, not because teachers are untrustworthy, but because they are human. The choice of a particular moderation method should recognise that distinction.

Methods of moderation

68. Methods of moderation have twin functions: to communicate general standards to individual assessors and to control aberrations from general standards by appropriate adjustments. Three methods are in common use:-

- scaling on a reference test;
- inspection by visiting moderators; and
- group moderation bringing teachers together to discuss their assessments.

Each differs in its relative emphasis on communication and control.

69. A reference test is any external instrument which has been set to the whole population of pupils who are assessed by their teachers. Scaling assumes that performance on the test is a better guide to the average and spread of the performances in a class than are the teacher's assessments and so the teacher's marks are adjusted - scaled - to bring them into line with the marks on the reference test. Usually only the rank order remains sacrosanct. The validity of the technique depends upon the reference test reflecting precisely the same features as the teacher's ratings.

70. The standardised assessment tasks of the national assessment system could be used as reference tests against which to scale teachers' ratings. This method would be economical, but there would be a price to be paid. For clear communication to teachers and parents about the attainment of the targets, the marks have to be meaningful, and mark ranges should be therefore be short. Thus it may be educationally desirable to use a reporting scale with only four levels, but any statistical adjustment then involves switching between two adjacent levels out of the four (for example, regrading all level 1 results into level 2 see paragraphs 97ff), which would usually be too gross an adjustment. On the other hand, the statistics of scaling require the teacher to use long mark ranges so that sensible timescale mark adjustments can be made. But this is in conflict with the need to communicate in terms of broad levels of performance for which long mark ranges are not necessary. Another disadvantage of scaling on a reference test is that because the test alone determines the level of reported achievement, teaching becomes dominated by test technique.

71. Moderation by inspection avoids this problem, but still has an overt emphasis on control, rendering communication one-sided. Furthermore, the control function itself is not efficient because it must suffer from incompleteness of data: only the final outcomes of tasks can be retained for inspection. Special arrangements may be made for the visiting moderator to observe processes and skills in operation, but these occasions are bound to be artificial and to become in effect an external practical examination.

72. In group moderation the emphasis is on communication through discussion and exchange of samples at a meeting. It is the favoured approach when examination boards scrutinise their own standards. Group moderation allows examiners to clarify in discussion both the objectives of their syllabuses and the bases of their value judgements. It is the only one of the three moderation methods which would enable the professional judgements of teachers to inform the development of the national curriculum. At the moment it is time-consuming and therefore expensive but it will be less so when assessment is based on explicit criteria.

73. Control can be built into group moderation. Teachers could bring to a group meeting two sets of results for each profile component. The first would be their own ratings; the second would be the results from the national tests. Both would be expressed as the distribution of their pupil groups over the levels of the national assessment scheme. A first task of a moderation group would be to examine how well the patterns of the two matched for each group of pupils. To take a simple example, if the results for a profile component were to be in terms of percentages of pupils in each of levels 1, 2 or 3, then teacher ratings might give 25% of their pupils at level 1, 50% at level 2, and 25% at level 3. On the other hand the national test results might for the same group of pupils indicate 20%, 60% and 20% respectively. The meeting would also have available the distributions over the whole group, both of the teacher ratings and of the results of the national tests (standard assessment tasks). They might also have results from all groups in an LEA; eventually, values derived from national norms may also be used, whether obtained in national trials or from results of previous years.

How group moderation works

74. The group moderation meeting would first explore any general lack of matching between the overall teacher rating distribution and the overall distribution of results on the national tests. The general aim would be to adjust the overall teacher rating results to match the overall results of the national tests; if the group were to have clear and agreed reasons for not doing this, these should be reported to the LEA and eventually to the body responsible for operating the national assessment system. Such reports should lead to a new scrutiny for possible defects in the tests if enough agreed discrepancies were to be found.

75. The meeting could then go on to explore discrepancies, in the pattern of the results of particular schools or groups, using samples of pupils' work and knowledge of the circumstances of schools. The moderation group's aim here would be to arrive at a final distribution for each school or pupil group. In general this would be the distribution on the national tests, with the implication that teachers' ratings would need adjustment, but departures from this could be approved if the group as a whole could be convinced that they

were justified in particular cases. The same pattern-matching technique could be used by LEAs so that they could check that local group practice did not begin to diverge from national expectations, without the need for a cumbersome hierarchy of subordinate groups based on representatives of local groups. It should be noted that a distribution such as 20%/60%/20% over three levels is more informative than an arithmetic mean. Distributions of 20%/60%/20% and of 10%/80%/10% could give the same overall mean score but for most purposes they are significantly different from one another.

76. The controlling element in this method of moderation can be achieved without a group meeting. The value of group moderation, however, lies in communication. It affords the opportunity for teachers to discuss possible interpretations of the most frequently occurring pupil profiles and their formative implications. Mismatches between internal and external assessments can be discussed in terms of the interpretation of the national curriculum and the effectiveness or otherwise of national assessment instruments. These professional deliberations have a valuable staff development function while at the same time, with appropriate documentation of decisions, providing formative feedback to the relevant subject groups concerned with the development of the national curriculum. The national assessment system is a communication system and professional communication should be its central core. A detailed account of a moderation meeting, based on experience of such work in the GCE system, is given in Appendix H.

77. It may not be possible to undertake group moderation for all profile components every year. In that case it could be undertaken in alternate years and experience may show that it is more essential, or more professionally valuable, for some components than for others. Where moderation is not undertaken through group meetings, it should proceed by simply adjusting the distributions to agree with those of the national testing.

We recommend that group moderation be an integral part of the national assessment system. We recommend that it be used to produce the agreed combination of moderated teachers' ratings and the results of the national tests.

Supplementary assessment

78. At this stage of the assessment process, a means has to be found of resolving the minority of individual cases where the internal and external profiles do not agree, and of confirming tentative diagnoses of serious pupil difficulties.

We recommend that further assessment instruments be available for teachers to use in cases where they require additional evidence about particular pupils.

79. Such further assessment instruments would serve three purposes. Some, specially developed, should be narrowly focussed on particular learning problems to confirm tentative diagnoses. The emphasis would be on error analysis but outcomes should be related to the same scale points as teacher ratings. The other two purposes could be met without additional test construction. To resolve discrepancies, national assessment instruments from previous years could be used. For pupils whose initial performance is out of the normal range for the age group, national assessment instruments from adjacent ages could be used. Most pupils would not be involved in this phase so demands on the teacher's time should be manageable. None of the instruments used for these three purposes would be the same as the more detailed diagnostic tests needed for pupils with particularly unusual difficulties.

80. Group moderation, with follow-up observations in a sequential and 'tailored' approach to assessment, would meet the objective of sound formative reports from teachers about, and to, individual pupils and to their parents, and from teacher to teacher. The soundness of this approach would be underpinned by the professional responsibility of teachers for their individual pupils.

We recommend that the final reports on individual pupils to their parents should be the responsibility of the teacher, supported by standardised assessment tasks and group moderation.

Aggregation for different reporting purposes

81. Confidence in reported results increases as the number of measures upon which they are based increases. Each of the purposes of the national assessment system can be met by a separate report based on a different aggregate to which a different degree of confidence can be attached. These aggregates can be at class, school, LEA and national level.

82. Formative assessments would take the form of a pupil profile, with each profile component based on a teacher rating and combined with items of information from the standard assessment tasks. A pupil's profile may change over time but this is not a cause for concern. It is indeed entirely proper: the very intention of formative assessment is to provide the information which teachers need to bring about change in the pupil's achievements.

83. The evaluative purpose of the national assessment system would be met by aggregations over large numbers of pupils. Sufficient confidence would attach to these aggregations, target by target, to provide formative feedback both to inform schools and others about their work and to guide those responsible for developing the national curriculum. When targets are further aggregated into profile components there should be a high degree of confidence in conclusions about the delivery of the national curriculum in most schools.

84. These reports are likely to be too complex to be used in facile comparisons of school performance. If, however, profile components are reported on a common scale, the aggregation over pupils and over profile components will yield data which might be used by parents to evaluate schools. Because of the aggregation there can be confidence in the measurement precision of this report. But the aggregation also means that only a very general interpretation of these results is possible. The measurement precision does however have an advantage: the distribution of achievement in a school should be interpretable with some confidence, broadening the basis of evaluation from the current narrow emphasis on average achievement.

We recommend that wherever schools use national assessment results in reports for evaluative purposes, they should report the distribution of pupil achievements.

85. The issue of reporting results at school level is further discussed in Section XII.

Cost effectiveness

86. Our proposals for the national assessment system have been strongly influenced by the importance of securing confidence in the resulting reports. At the same time they are compatible with considerations of economy.

87. The heaviest reporting load would fall upon class teachers of the 11 year old group, who may have to report on more than a dozen profile components per pupil. Such teachers already carry a heavy responsibility in communicating pupil performance to the receiving secondary school. The more structured approach of the national assessment scheme, with its greater degree of external support, should go some way to offset the increased reporting load, and absorb much of the effort currently being taken up by unsystematic assessment.

88. Much of the assessment activity would develop from existing good practice. Teachers' ratings of profile component performance would be based upon whatever internal evidence is currently used by the school when reporting to parents, though with the greater clarity stemming from defined attainment targets.

89. The externally provided assessment tasks would, at the reporting ages, supplement and in part substitute for classroom tests designed by the teacher. The tasks should help to disseminate a sound methodology for probing pupils' achievements for formative purposes. They should also help to communicate national achievement data for the age group, and a range of differentiated targets above and below these expectations.

90. The sequential assessment strategy described above, coupled with a staged progression of targets, would simplify the design of the tests. Each could be focused on a single stage and probe half a dozen profile components. Designed as normal curriculum material, they could be distributed across a term's work with little additional demand on pupils' time. These externally devised assessment tasks, although expensive to produce initially, would have a utility beyond the reporting ages in a staged system. Teachers could use them when

appropriate in intervening years to support their judgements of progress conveyed at meetings with parents. In these ways we envisage that the national assessment system will support teachers in every year of schooling, not just the four reporting years.

THE ASSESSMENT SYSTEM IN PRACTICE

XI PROGRESSION, SCALES AND AGES

91. As emphasised in our Introduction (Paragraph 5) we regard it as one of our priorities to ensure that criteria and scales used should relate to expected routes of development, giving continuity to each pupil's assessment at different ages, and thereby giving a detailed picture of each pupil's progress.

Ages of assessment and reporting

92. To provide a clear basis for subsequent discussion, we first consider the ages at which national assessment and reporting should take place. For national purposes we believe that the balance of advantage lies with reporting towards the end of the school year in which a cohort of pupils becomes 7, 11, 14 or 16. With very few exceptions, the ages of 7 and 11 come either at the end of or, in middle school systems, at least two years into a phase of schooling. By those ages children should therefore feel settled in that phase, and the teachers should know them well. It is important that 7 year olds who have not made a satisfactory beginning in learning to read, write and calculate should be identified and helped to make progress. Where there is transfer, whether after 7 or after 11, the process of reporting should allow useful information to be passed to the next school. The age of 14 is desirable because it is commonly the time when decisions are being taken about the courses children should subsequently follow; and 16 allows a summative assessment to be made for those children moving out of the school system, whether to work or to further education or training.

We recommend that the ages for national assessment should be 7, 11, 14 and 16, with reporting occurring near the end of the school year in which each cohort reaches the age involved.

Progression

93. Given a framework of reporting at those four ages, we go on to assume progress to be defined in terms of the national curriculum, and the stages of progress to be marked by levels of achievement as derived from that curriculum. It is not necessary to presume that the progression defined indicates some inescapable order in the way children learn, or some sequence of difficulty inherent in the material to be learnt. Both of those factors may apply, but the sequence of learning may also be the result of choices for whatever reason, which those formulating and operating the curriculum may recommend in the light of teaching experience.

94. The consultative document on the national curriculum, published in July 1987, proposed that 'the main purpose of assessment will be to show what a pupil has learnt and mastered ...' (paragraph 28); that 'attainment targets will be set' for the core subjects, establishing what children should normally be expected to know, understand and be able to do at around the ages of 7, 11, 14 and 16, so enabling the progress of each child to be measured against established national standards (paragraph 23). The subject working groups should, among other things, take account of the need for continuity and progression throughout compulsory schooling (Annex A, 3.1.c).

95. The profile components described in section VI, as defined by the subject working groups, will identify the main areas of attainment within which progression is to be assessed. In the following paragraphs we consider how information about the pupil's stage of learning can best be conveyed.

Communicating information about stages of learning

96. Most straightforwardly, one could simply describe what a pupil understands, knows and can do in terms similar to those that will be used by the subject working groups to define attainment targets. We believe that it would be right for much of the communication between teachers and parents to take that form at the reporting ages and between them. However, two considerations make it advantageous also to convert the descriptions into a marking scale. One is to convey some sense of where a child is in the process of learning; the other is

to make easier the analysis of results from groups of children. The last is of interest to parents who wish to know where their child stands in relation to others of about the same age and also, at various levels of aggregation, to teachers, heads, LEAs, the Government and the wider community.

97. Commonly used systems of scaling pupils' achievements divide those of about the same age into 5 or 6 groups. In England and Wales, some schools use a five point scale: A to E, for example. In the Federal German Republic a 6 point scale is common, and the numbers 1 to 6 are used. Generally the scaling is based on assessment, including the results of tests, which judge a pupil's performance against that of others in the group, or some comparable group: that is, the judgements are norm-referenced. Furthermore, the reference groups are usually of pupils of the same age group. Commonly, in the five-point system the scaling is so arranged that small proportions of pupils are likely to be in grade A and E, larger proportions in B and D, and perhaps half of the children in C. These proportions may vary when the groups being assessed are untypical of the general population. It is, of course, possible deliberately to arrange for the proportions to be different, and a five point scale could be adjusted, for example, to divide the pupils so that very few are judged to be 1 or 5 and the rest are spread evenly between 2, 3 and 4.

98. Levels of achievement could therefore be translated into a five- or six-point scale at each of the reporting ages. This could have the attraction of apparent uniformity across all ages. However, age-specific grading has drawbacks. Pupils are more likely than not to remain in the same grade or level, plus or minus one, at each reporting age. A child might be assessed to be on level 1 at one reporting age, on level 2 at the next and back to level 1 at the third, despite, in absolute terms, having made progress. Even if the child remains on the same level from age to age, he or she will be given no sense of having made the progress which must in fact have been made. Furthermore, even though level 3 at 7 years of age would indicate that a child had achieved specific competences, those same competences would merit a different level number for a slower child, only 1 or 2, at the next reporting age. There would be no simple relation between the level numbers and performance.

99. More generally, the combination of a norm-referenced system with age-specific scaling would not be consistent with the proposals in the national curriculum consultative document. The overall national purpose is to work for achievement of the attainment targets of the curriculum. Assessment, whether for feedback to pupils or overall reporting and monitoring, should therefore be related to this attainment ie it should be criterion-referenced. Given this, it follows that different pupils may satisfy a given criterion at different ages: to tie the criteria to particular ages only would risk either limiting the very able, or giving the least able no reward, or both.

100. In view of these disadvantages of age-specific scaling, we consider that the scaling system should be directly related to the development of pupils' competences as described by the attainment targets of the national curriculum. That is to say, scales should be used that indicate where a pupil has reached in a profile component. We shall use the word level to define one of a sequence of points on a scale to be used in describing the progress of attainment in the profile component. The sequence of levels represents the stages of progression.

101. A pupil assessed as achieving a given level, say level 2, will have satisfied the criteria for level 2 and will be working towards the criteria for level 3. Progress is marked by achievement of successive levels over time. For the purposes of national assessment, a broad indication of progress is required. We propose that the criteria defining successive levels be so chosen that a pupil could reasonably be expected to progress by one level in two years of work in that profile component. Over the age range 7-16 this would imply a need for 5 or 6 levels. At both ends of the age range, however, some will be unable to progress as fast as others, and some will be able to make quicker progress. This leads us to recommend that a total range of ten levels will need to be defined: fuller justification of this number will be given below.

We recommend that each of the subject working groups define a sequence of levels in each of its profile components, related to broad criteria for progression in that component. For a profile component which applies over the full age range 7 to 16, there should be ten such levels, with corresponding reduction for profile components which will apply over a smaller span of school years.

Progression through the levels

102. The implications of this proposal may be seen from the illustration in Figure 1. The details of the representation in this diagram will be explained in three steps below, considering respectively age 7, age 16 (with GCSE) and then intervening ages.

Age 7

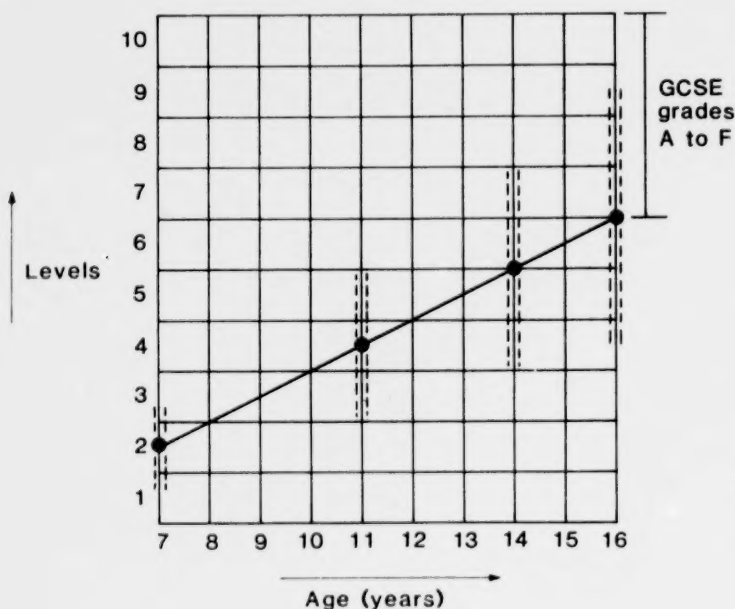
103. The first step in the explanation applies to age 7. The range of levels likely to be achieved by the youngest children will be narrower than the range likely to be achieved by the oldest. For reasons that are outlined in section XIII, we believe this to be advantageous, and we recommend that, for reporting purposes, the levels for the 7 year olds should be confined to levels 1, 2 and 3. These levels should, for this age group of children, be regarded as mainly diagnostic in the sense that level 1 identifies those children who may need more help in their learning than can be provided from ordinary classroom resources; those in level 3 have made such advances that they also need additional help if they are to maintain their speedier progress. We recognise that assessment at levels 1 and 3 will be more common in some aspects of learning than others; for example, there are likely to be more children in level 1 for literacy and mathematics than for some other aspects of learning.

We recommend that levels 1 to 3 be used for national assessments at age 7.

Age 16

104. To develop the second step of the explanation of Figure 1, we note that at the other end of the age range, there has in the past been a variety of attempts to derive written criteria from GCE and CSE grades. That experience suggests that no more than four sets of levels can be identified in relation to the GCSE population.

FIGURE 1 SEQUENCE OF PUPIL ACHIEVEMENT OF LEVELS BETWEEN AGES 7 AND 16



The bold line gives the expected results for pupils at the ages specified. The dotted lines represent a rough speculation about the limits within which about 80% of the pupils may be found to lie.

105. We are conscious of the tendency for new assessment systems to overlay rather than to replace existing schemes, resulting in a growing burden on the schools. It would be unfortunate to have two independent systems in operation for the 16 year old age group. On the other hand, problems of comparability are not readily solved and a premature attempt to link the national assessment system with GCSE could lead to a top-down approach which placed undue constraints on the design of the national curriculum at earlier ages. It is our expectation that the phasing of the introduction of national assessment (see Section XVII) will work towards GCSE rather than away from it.

We recommend that the formal relationship between national assessment and GCSE should be limited, in the first instance, to this one reference point: the boundary between levels 6 and 7 should correspond to the grade F/G boundary for GCSE.

106. The four levels from 7 to 10 could, following this recommendation, bear some relationship to upper GCSE grades, but we believe it would not be helpful at this stage to attempt to specify equivalences between the two systems whilst they operate in different ways. These upper levels should be set by criteria which specify attainment targets appropriate to the GCSE population at age 16, although again this relationship cannot be precise.

We recommend that as they develop the upper four levels of their profile components, the subject working groups adopt present practices for determining GCSE grades at A/B, C/D, mid-E, and F/G as a starting point.

107. It has been suggested by some that GCSE should be revised and extended to provide appropriate targets for the entire age group. We feel this would be unwise on three grounds. GCSE is a major educational innovation which has led to considerable upheaval in the secondary school system: teachers need a period of stability to come to terms with these changes. Secondly, if the national assessment system meets its objectives, a larger proportion of the age group will be served by the existing GCSE system. There is value in setting up a dialogue between the GCSE system and the curricular approaches characteristic of earlier years of education and the time scale allows for this. Finally, in a system of progression, extension of GCSE in the short term is unnecessary, since the levels designed for the middle range of 14 year olds will present appropriate challenges for pupils who at age 16 have not yet progressed into the GCSE grade range.

We recommend that GCSE be retained in its present form until the national assessment system is initiated at earlier ages.

Ages 11 and 14

108. To develop the third step in the explanation, we return to the assumption that a level should be roughly equivalent to two years of educational progress in a profile component. The average expectation for 14 year olds will therefore be the level 5/6 boundary. On similar grounds the average expectation for an age 11 pupil will be level 4. In mathematics, the widely recognised seven year gap in attainment between the most advanced and the slowest pupils at age 11 would imply that levels 3, 4 and 5 should be characteristic of profile components in that subject for the age group but that there will be a few pupils who have already progressed to level 6, the highest level of report for junior school pupils. In general, however, level 6, which should give an indication of how levels might be interpreted in relation to GCSE, should be achieved by some pupils at 14. Level 6 at age 14 is indicative of a pupil who is on target for a GCSE grade. At earlier ages, however, not all of the requisite knowledge and experience may be in place. Furthermore, GCSE grades involve combinations of profile components, and some components may be encountered only late in the curricular experience of the pupil.

109. In reviewing the above argument, it should be emphasised that the basis is the definition, by subject working groups, of a sequence of levels so tied to criteria that two years' learning represents one level of progress. The spread that may be found at any one age, given such a system, remains to be discovered, but the above speculations have some basis in practical experience. For instance, a growing body of evidence indicates that in English and mathematics there is a very broad range of achievement in any year group, sufficiently broad to ensure that there will be overlapping achievements at ages 7 and 11, 11 and 14, 14 and 16. Thus a curriculum target designed for one age group may be quite appropriate for some pupils in an adjacent age group. The need for differentiation at any particular age, coupled with the overlap of achievements between reporting ages, supports our proposal for a single sequence of levels across the age range for national assessments. Indeed in our view the differentiation of attainment targets at a particular age arises naturally from the notion of age progression upon which the reporting scales of national assessment are based. In a criterion-referenced system there should be no distinction between the definition of the sequence of levels of a profile

component to reflect progress between ages, and its use to differentiate progression at a particular age. Only one set of criteria is required. The levels defined by the national curriculum attainment targets will provide differentiated challenges at each age according to the needs of the individual pupil. All pupils should then have an expectation of making progress in every profile component: progress which will be indicated by the achievement of new targets.

Differentiation

110. This leads naturally to a discussion of differentiation. As we shall argue in a later section, we think that for the youngest age tests should be devised that, from the children's point of view, are similar in form to their normal school work, should be based on familiar material, and should allow children to exhibit different levels of knowledge, understanding, and skills in their responses. So as to take account of the varying maturity of pupils and the burden of testing on the class teachers, considerable restraint in the number of assessment tasks will also have to be exercised at the 7 and at the 11 year old stage, and this is further discussed in Section XIII.

111. At the later reporting ages it may not always be possible to rely on differentiation by outcome in this way, and the tests devised may themselves have to be sequenced according to the targets being tested, though within the same profile component.

112. Given the view that the differences in performance at any one age must be appraised by the same criteria as progress between ages, it follows that any pupil who has made remarkably fast or remarkably slow progress may well be assessed by the methods used for normal pupils at much older or much younger ages respectively. Thus it would not be impossible for a pupil at age 7 to be capable of level 4 in some components. With this strategy, which is easy to implement because of the cross-age sequence of levels, there would be no need, for the purposes of reporting within the national assessment system, for pupils to take tests or other assessments at earlier or later ages than those specified.

We recommend that assessment and reporting for the national assessment system be at the same ages for all pupils, and that differentiation be based on the use of the single sequence of levels set up to cover progression over the full age range.

Improvement over time

113. The normal distributions across levels at each age will indicate that the attainment targets corresponding to those levels constitute the range of targets appropriate to that age. But as curriculum and teaching are adapted to the new system, improvement in these distributions may be expected: these can be recorded and investigated in our scheme because it is criterion-referenced and not tied to limited expectations at particular ages. Nevertheless, for the reporting and subsequent analysis to be effective, it is important that virtually all pupils be assessed at the same age. It would be possible in any analysis to take account of the fact that (say) 2% of pupils at age 14 achieved level 9 in a particular profile component (and the means for them to exhibit this would be available); it would not be possible to feed into analysis the fact that they were allowed to attempt assessments for level 7 at age 12 and, having succeeded, were excused at age 14.

114. For reporting at a particular age, it will be possible to say of pupils in a particular class, or school, or LEA, what proportions were in each of the levels of the sequence. As recommended above, that set of proportions would constitute the form of report for each profile component.

Special needs

115. In considering the guidance appropriate for individual pupils, the teacher should be giving particular attention to cases of special need. For example, any children who do not move into a higher level between reporting ages should certainly have been identified as needing extra help, as may some who move through the level system slowly. There may be particular difficulties with pupils who might not be able to undertake the normal assessment tasks. All children who are assessed in the agreed profile components should take part in the nationally prescribed tests unless the headteacher certifies that a child

with special educational needs is, by reason of those needs, unable to participate in the test. Some children may be able to take tests in an adapted form or with special help - which should be taken into account in determining the results (see Section XV). Children who are statemented will be exempt if their statements exclude their participation in the relevant aspect of the national curriculum. (See Section XV for further discussion of special needs.)

Continuing assessment between reporting ages

116. Whilst most of the above discussion has concentrated on assessment and testing at the four specified ages, it will be necessary to set this programme in the context of the frequent assessment of pupils' work that teachers make, and will continue to make. This work is far more extensive than that of the national assessments, covering progress over all of the periods intervening between 7, 11, 14 and 16. The system of levels tied to criteria which we propose should provide a valuable framework for relating this work to the attainment targets of the national curriculum. A training programme should be arranged to help teachers to become more effective in the frequent assessments that they make, and to help relate these to the criteria embodied in the national tests and assessments. This need was envisaged in paragraph 74 of the consultation document on the national curriculum.

We recommend that support items, procedures and training be provided to help teachers relate their own assessments to the targets and assessment criteria of the national curriculum.

117. Teachers will also need a varied set of diagnostic tests to help explore special difficulties shown up in the national and other assessments.

We recommend that a review be made of the materials available to schools for detailed diagnostic investigation of pupils' learning problems, and that the need for extra help with production or advice about such materials should be considered.

These recommendations should be considered with the complementary recommendation made in paragraph 78.

Assessment of new profile components

118. Most of the foregoing discussion has been set out on the assumption that it is about profile components that are applicable across the full age range from 7 to 16. We recognise, however, that it may be necessary, in some aspects of the curriculum, to introduce new profile components as the children progress, and that these new components may be suitable only for children who have already acquired certain levels of knowledge and understanding. It is for the subject working groups to recommend whether and about what, this should be so. The initial level for these wholly 'new' profile elements should be level 1. A modern foreign language is a special case, since under the national curriculum this would not normally start being taught until the secondary school stage; the reporting system for all profile components in a foreign language will need to begin at level 1. We regard this clear divorce of level number from age to be advantageous rather than otherwise, since it will emphasise the criterion-referenced nature of the assessment.

Number of profile components assessed at primary stage

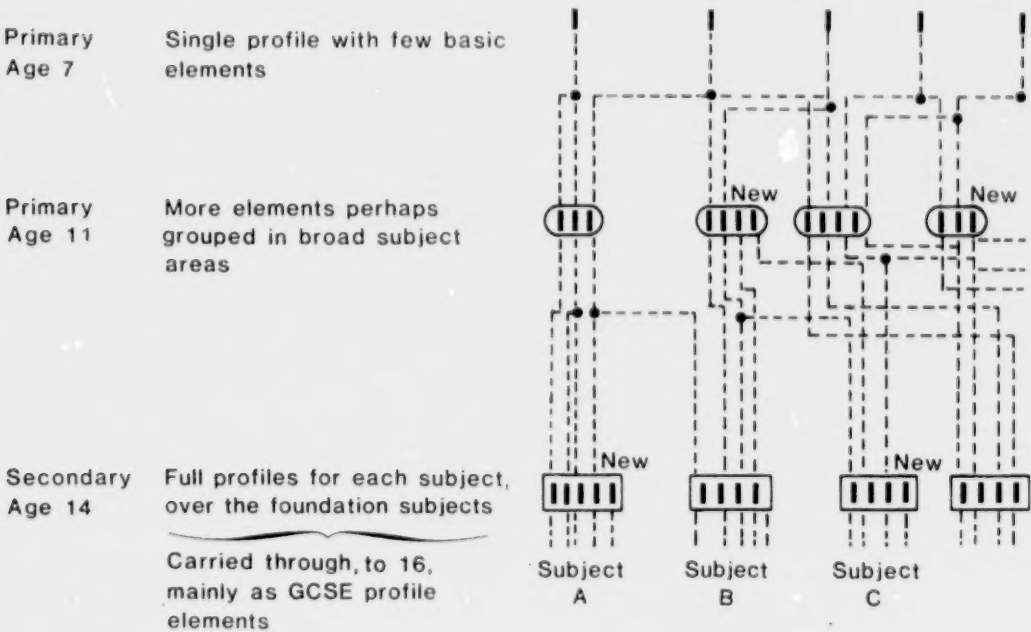
119. Whilst it follows from the above that the number of profile components to be assessed at 7 and 11 will be less than at 14 and 16, there is still a potentially serious problem because of the size of the burden that could too easily be placed on teachers in the primary phase. If the subject working groups were to work independently in making recommendations for primary pupils, then a single class teacher of 7 or 11 year olds could have to assess pupils in at least 20 and possibly 30 or more profile components. Furthermore, some components of foundation subjects would be inappropriate, in their fully differentiated form, at least at age 7.

120. One solution would be to limit assessment in the primary phase to the three core subjects. This could have the effect of narrowing the primary curriculum to an undesirable extent, even if it were only adopted for age 7.

121. There is a different approach to this problem which could be more fruitful. As recommended earlier, subject working groups should pay particular attention to profile components which might overlap, or be identical, between different

subjects because they represent important cross-curricular areas of knowledge understanding and skill. We expect that some profile components will have a wide range of applicability across the subjects of the curriculum, so as to encourage pupils, teachers and parents to recognise common and overlapping aspects in the work being reported. In addition some profile components which should be separately identified in separate subjects in the secondary phase could well be combined up to age 11 because the aspects which differentiate them are not relevant at younger ages. Thus profile components at older ages might evolve by differentiation from a smaller number at younger ages. Figure 2 illustrates this approach.

FIGURE 2: EVOLUTION OF PROFILE COMPONENTS



122. Whilst this approach reduces the numbers of profile components at 7 and 11, it does not neglect the emergence of subject identification at age 11, indeed it allows the context in which assessments and tests are set to become increasingly identified with curriculum subjects as pupils mature. One most important advantage of the scheme represented by Figure 2 is that it stresses the need for continuity in assessment, and therefore in curriculum planning, across the whole age range from 5 to 16. Thus, work at age 11 should both have a planned relationship to the work suitable for 7 years olds and also provide a suitable basis for more highly differentiated work in the next phase. The information provided from national assessment at 11 would then be relevant to the interests of all secondary subject teachers in appraising the starting points for work with pupils after age 11.

123. It is not possible for us to develop this theme beyond the very general form sketched in Figure 2. The underlying issue is a curricular one and it is for the subject working groups to collaborate with one another in their recommendations for 7 and 11.

We therefore recommend that a working group be established, with some shared membership between the subject working groups, to co-ordinate their proposals for assessment, including testing, at the primary stages, in the light of a comprehensive view of the primary curriculum and of the need to limit the assessment burden on teachers.

124. In summary, progression would be identified through the reporting system on the basis of identifying changes in what the children know. Many of those changes would be denoted by the changes of level within a profile component. Some components would be subdivided as learning progresses, and the level numbering would continue. Some might be wholly new and the level system would commence from level 1. New components and some sub-divisions of components may be for use only with children who have reached pre-defined levels or, in the case of a modern foreign language, the appropriate age. At the first reporting age, some profile components would be general and not specific to subjects, but would form the basis for a growing number of subject components which would be developed from them at later ages. Thus progression would be firmly emphasised across the transition from primary to secondary work.

XII REPORTING

125. The results of national assessments will need to be reported to a variety of audiences; the intended audience will determine the degree of detail and interpretation required.

Parents and individual pupils

126. Parents require information about their child's progress in relation to the attainment targets of the national curriculum in a variety of areas and presented in such a way that the individual child's progress can be related to national performance. It is important that results communicated to parents and children should be positive - indicating what the child has achieved; and that they should be constructive - pointing to what needs to be done if the strengths and weaknesses identified by assessment are to be dealt with appropriately. Assessment has to be designed to help all children to proceed with their schooling to the limits of their ability; weaknesses have to be recognised, and appropriate help made available.

127. Thus reports to a pupil and to the pupil's parents should be in terms of profile components since these give the necessary detail about strengths and weaknesses in order to guide future work. As Records of Achievement are more widely implemented, the profile component results should be incorporated into the record, which will also contain other types of information about the pupil (see Appendix I).

Reporting within the school

128. It should be for the teachers within a school to decide how profile component data should be aggregated to provide a basis for internal review. Data aggregated at class level over separate profile components might be of value to a teacher co-ordinating work in an aspect of the curriculum in a primary school, or within a subject department in a secondary school, whilst aggregation over components to give single subject scores might be appropriate in reviews of the school as a whole. Data should show the distributions of performance rather than a single mean and could be compared with the overall national performance.

Employers and further education

129. Results of national assessment of individuals at age 16 may also be taken into account by, for example, employers making decisions about offers of employment, apprenticeships and YTS places and, among others, sixth form and FE colleges making decisions about suitable courses. If the results of assessment at age 16 were included in the pupil's Record of Achievement, employers and colleges would receive a wider view of a young person's potential than could be provided by the results of the national assessment system alone.

Educational decision-makers

130. School governors and local and national providers of education will also need aggregated information about the distributions of performances, within classes and within different subjects. Local and national authorities would have to specify the forms and levels of aggregation that they require. For these audiences scaling or adjustment of distributions should not take place. Parents should also have access to results for the overall performance of the school which their children attend.

Confidentiality for individual pupils

131. If national assessment scores for a school are to be reported to a wider audience than to individual parents, governors and providing authorities, then several problems will have to be considered. It should first be clear that results for individual pupils are confidential, to be shared with each pupil and with his or her parents, and to be transmitted within the school, to a receiving school on transfer, and, if necessary, to the local authority, only in confidence.

We recommend that national assessment results for any individual pupil be confidential, to be discussed between pupil, parents and teachers, and to be otherwise transmitted in confidence.

National assessment results for a class as a whole and a school as a whole should be available to the parents of its pupils.

Reporting a school's performance

132. The general problem is that a school's "performance" can only be fairly judged by taking account of many aspects of its work and of many factors outside its control that affect its work. Performance in the national assessment will only be one element in this broad picture. So any report of assessment results should be part of a broader report covering many activities and achievements of the school, including information that the school may or should publish about its curriculum.

We recommend that the only form in which results of national assessment for, and identifying, a given school should be published is as part of a broader report by that school of its work as a whole.

133. It is well known that a variety of socio-economic factors so affect pupils' capacity to respond to school work that they can lead to very wide differences between the performances of different schools (see Appendix J for a review of the evidence). It is difficult to see how such factors can be taken into account in reports of the performance of individual schools. A number of issues arise. Firstly, should the new data be scaled to account for such factors before being presented? Adjusted figures, by scaling up the scores of pupils with low attainments in areas of socio-economic deprivation, could conceal large numbers of pupils who are under-achievers and could also conceal the need for additional resources to be made available to these schools. On the other hand, a misleading view of a school's performance could easily result from the use of unadjusted data as a basis for judgement. For example, average attainments from pupils drawn from a favoured area might appear satisfactory but would in fact indicate that the school was under-performing in relation to the capacities of its pupils. We take the view that the publication of school results to a wider audience than parents, governors and providers and the use of such results to compare schools' performance would be liable to lead to complacency if results were adjusted, and to misinterpretation if they were not.

134. Reports to the public of national assessment results should therefore always be by means of a report produced by the school, authenticated by the local education authority in the case of those schools for which that authority

is responsible. Results should not be adjusted for socio-economic background, but the report should include a general statement, prepared by the Local Authority for the area, of the nature and possible size of socio-economic and other influences which could affect schools in the area.

We recommend that any report by a school which includes national assessment results should include a general report for the area, prepared by the local authority, to indicate the nature of socio-economic and other influences which are known to affect schools. This report should give a general indication of the known effects of such influences on performance.

135. We are agreed that the age 7 results of national assessment would not be a useful guide to the quality of education in a school, because of the large and variable effects both of home background factors and of differences in entry ages and pre-entry experience of pupils. This latter factor is discussed in more detail in the next Section. We therefore believe that there should be no requirement on a school for publication of the school's age 7 results to a wider audience than parents, governors and providers. Nevertheless, if in spite of the reasons advanced above, an individual school wished to proceed with the publication, then it should be allowed to do so at its own discretion.

136. A characteristic of primary schools, as distinct from their pupils, that needs to be taken into account in settling on the reporting system for the national curriculum is their size. Most primary schools are small as compared with secondary schools. (See Appendix K on pupil numbers). In particular, the number of pupils in an age group in most primary schools is below 30, and in a significant proportion of schools it is below 10. The number is frequently too small to allow any general conclusions about a school to be drawn from the levels of achievement reached by a single age group, and it may be necessary to consider results over a number of years. That also presents difficulties. Over even as few as 5 years, the staff of a school might have changed considerably and so, probably more often in inner city than in rural areas, may the nature of its catchment area. The aggregated results need to be interpreted with great care.

137. For primary schools with age 11 pupils, we consider that data on national assessment at age 11 should be provided only in the context of a school report as set out above; if schools so desire, data may also be provided on the same cohort at age 7, if these were in the same school at that younger age. These reports would offer general guidance to readers about catchment area effects and be available to any enquirer, although written particularly to guide prospective parents.

We recommend that national assessment results, for pupils at age 11, aggregated at school level, should be published as part of each primary school's report. We recommend that there should be no requirement to publish results for pupils at age 7.

138. For a secondary school, the publication should include its GCSE and other national assessment results for age 16; its national assessment results for age 14 (in both cases showing the distribution of such results); and, if the school so desires, information about progress made by the 16 year old cohort in relation to its national assessment results at age 14. A school with age 11 entry might also, at its discretion, refer to overall data for the distribution of national assessment results for its entry, but this should be done without exposing data on individual feeder schools.

We recommend that national assessment results for pupils at ages 14 and 16, aggregated at school level, should be published as part of each school's report.

XIII THE PRIMARY STAGE

139. We turn now to the implications for work in primary schools of the recommendations we have made.

140. The idea that children's achievements should be described according to what they know and can do is central to the recommendations made in this report. The assessment programme depends upon a clear view of what children should be being taught and should learn: the curriculum. We are not satisfied that a sufficiently clear and coherent view of the primary school curriculum now exists

in relation to the national curriculum, and the subject working groups will need to work in close harmony if they are, between them, to define such a curriculum with its associated attainment targets.

Training

141. The sizes of primary schools and the educational immaturity of the pupils have consequences for the ways in which they are organised. Primary school teachers are more often than formerly using their special interests and skills to advise their colleagues and to teach children in addition to those in their own class. However, class teaching remains, and will remain, predominant. From year to year teachers do and should move from teaching one age group to teaching another. Indeed, in a substantial number of classes there are children of more than one age group.

142. At least two things follow. All primary school teachers will need to be trained in assessing at least one of the first two reporting ages, and many may need to be trained for both primary ages; moreover the teachers will need additional training in assessment across the full curricular range of the assessment programme. The arrangements for training primary school teachers with regard to the national curriculum will have to take account both of the large numbers of teachers to be trained, and of the considerable load on each if everyone is to operate without easily accessible support in all areas of the curriculum.

Literacy and numeracy

143. We make the assumption that assessment in primary schools must be concerned, inter alia, with the development of children's powers of literacy and numeracy, the first of which has applications well beyond what might be contained within English, and the second of which should certainly not be the limit of work in mathematics. There are many skills and concepts additional to these that are essential to children's learning both in the primary school and later. It is not for us to categorise them, but we recognise their importance.

Some, and perhaps many, are of general application, in the sense that they are necessary in the study of a number of subjects. In expressing such judgements, we recognise that we are making assumptions about an overall view of the primary curriculum that might inform the strategy for assessment and testing.

Organisational factors

144. Most children are taught in school systems that arrange transfer from primary to secondary schools at about 11 years of age. Some children move from infant to junior school at 7 and others attend a full-range primary school from 5, or earlier, until 11. There are also various systems with middle schools. In some, children attend combined first and middle schools until 12; in others they change school at 8 or 9 (or 10 in a handful of cases) and then, respectively, at 12 or 13. In some areas there is transfer at 14. Appendix K on school types and numbers gives detailed statistics.

145. We prefer the reporting ages in the primary stage to be such that the children will have had at least 2 years to settle in their schools and be well known by their teachers. We also regard it as important that sufficient information should be available to a receiving teacher or school so that it is possible for that teacher to provide a programme that takes children forward, rather than to spend time attempting to discover what children know. On balance, therefore, and bearing in mind the preponderance of systems where the transfer age to secondary schools is 11, we have already proposed that the first two reporting ages for national curriculum purposes should be the school years in which children become 7 and 11 years of age.

146. At 7, a number of factors have to be kept in mind. The children may or may not have had pre-school experience in a mother and toddler club, in a pre-school playgroup, in a social services day centre with child-minders, or in a nursery school or class. Where children are admitted to school at and not before the statutory age, some 7 year olds have had 3 years in school and some only 2. In a growing number of areas, children are admitted before the statutory age: they may be admitted at the beginning of the school year in which they will be 5; or at the beginning of the term in which they will be 5; or at some intermediate

time. In areas where social behaviours in the children's homes are different from those they find in school, the children have had less time than at 11 to adjust to school requirements. Those from homes where English is not the first language have had less time than at 11 to acquire fluency in English. The reporting age covers children with an age range of a year: the age difference is proportionately greater at 7 than at 11.

Assessment at age 7

147. We conclude that national assessment at the first stage should be aimed principally at identifying children in need of help which is exceptional for a child of that age, whether because he or she is making so little or so much progress in a profile component. Such help may in practice be provided either by the class teacher or in other ways. There is no suggestion here that all such children are candidates for statementing as defined by the 1981 Education Act, or that special help with regard to one profile component implies that special help is required with respect to other profile components.

148. The results of assessments should be recorded on the pupils' school records and could be of value in reviewing subsequent progress. Care should be taken at all reporting ages, but especially at this age, to avoid giving the impression that the assessment is a prediction of future performance.

149. Because of the immaturity of the 7 year old children and the practices ordinarily found in schools, we consider that externally provided tasks should be no more than three in number for each child. They should seem to the children to be part of ordinary school work, though conducted in a standardised way by the teacher. The tasks should not be differentiated according to levels of difficulty, but should be so constructed that children's performances on them can be related to the national curriculum attainment targets. The topics should be selected by the teacher from item banks so that the matter on which the tasks are based will be within the children's general experience. The three tasks should be drawn from an item bank so as to allow the children to demonstrate clearly the stages reached in literacy and numeracy and the range of other skills deemed appropriate within the national curriculum. Appendix D gives

examples of such tasks: each illustrates how in the course of work that should engage pupils' interests, opportunities arise for assessing the competence of pupils in sharply defined profile components. Across a balanced set of three such tasks it will be possible to appraise all of the appropriate components on several occasions and in several different contexts.

150. It has to be remembered that a proportion of children, not least some with special educational needs, are not literate at 7. This should be allowed for in the standard procedures laid down for the assessment tasks. Children should not feel a sense of failure because they are unable to respond to a assessment/task, and it may be necessary to allow children who begin to falter to be helped under prescribed circumstances which, clearly, would lead to it being recorded that the help had been given. The requirements of moderation [see Section X] may well make it necessary that schools in a moderating group should agree on common use of one of the three assessment tasks to be used. We suggest that, during the early years of the national curriculum, common choice should be a requirement, to be abandoned if and when it is found that moderation can be effective even though different schools within a group choose no assessment tasks in common.

We recommend that at age 7 the standard assessment tasks for the national assessment should comprise a choice of three prescribed tasks for each child; each task should be designed to give opportunities for systematic assessment of competence in the range of the profile components appropriate to age 7.

151. As at other reporting stages, the results of the tests and the recorded continuous assessments made by the teacher should be considered together when deciding what level a child has reached.

Assessment at age 11

152. By the second age of reporting, that is at 11, the children are likely to be working in a mixture of topic and subject modes. The subject working parties should consider what applies in their field and recommend in what form profile components are presented. We anticipate that there will be requirements in a

subject context, and it is important to bear in mind that secondary schools, though they may undertake some aspects of their early work in a thematic form, may want information that can be linked to a subject analysis of the curriculum. We have already emphasised the need to strengthen collaboration between subject working groups to enable this approach to be made effective.

153. There could still be a particular burden upon the teachers of 11 year olds. Even if the assessment tasks were limited to three as for the 7 year olds, they would have to be more complex to permit the children to show the variety and depth of their competencies. The volume of work they would be likely to produce would be substantial and a thorough, standardised assessment of all its components would take a long time, especially when moderating procedures are allowed for. Without more advice from the subject working groups, we can only draw attention to the issue and tentatively suggest ways forward. It might be possible, for example, to concentrate the process of moderation each year on a sub-set of profile components across the assessment tasks or to limit moderation to two of the tasks. It might be possible to include some work within focussed tasks or tests that would be relatively highly standardised and thus simple to mark and requiring no moderation after marking. Perhaps the three tasks should focus respectively on mathematical and scientific learning; on literacy and the humanities, and on aesthetics. It may be that the number of tasks has to be increased to four. But we must stress that these groupings are primarily matters for the subject working groups; our examples are directed only at illustrating the nature of the issues.

We recommend that at age 11, the tests for national assessment should include three or four standard tasks which cover a range of profile components, possibly supplemented by more narrowly focussed tests for particular components.

154. Between each of the reporting stages, continuous assessment is a normal feature of classroom practices and this may be supplemented by diagnostic tests. As we have already noted there is a need for a wider range and variety of means of assessment to be developed for use in busy classrooms at the discretion of the teacher.

Summary of primary school procedures

155. In summary, national assessment in primary schools might work as follows (a fuller description is offered in Appendix L).

156. During the year in which the pupils become 7, teachers would be asked:

- on the basis of continuous assessment allied to programmes of study and attainment targets, to arrive at judgements of their pupils across a number of profile components and using three levels of performance;
- to select from national item banks three standard tasks from within the range of work covered, at least one of which would be chosen in common with other local schools;
- using the standard procedures and criteria for administration and for recording of performance, to incorporate the test tasks into their teaching programme so that each pupil is assessed, over three tasks, in the profile components;
- to attend a local group moderation meeting at which they consider together the results of the tests and their assessments, decide the overall distribution, resolve discrepancies and adjust individual levels in the light of the agreed distribution;
- to compile a confidential report for consideration by head and governors on the overall performance set against local and national distributions;
- to determine the levels for each pupil in the light of the profile components and of general information on standards as a whole, devised from the moderation meetings, and to report to parents;
- to enter results in the school records.

157. For pupils aged 11, the same procedures would apply, except that:

- there are likely to be three or four standardised tasks and these may be supplemented by other forms of test, externally provided, for particular profile components;
- the profile components would be more numerous and more complex;
- the information about overall school performance would be incorporated in a published report, related to a statement of the school's context;
- information on individual pupils would be made available in confidence to the receiving secondary school.

XIV THE SECONDARY STAGE

Assessment at ages 11-14

158. Assessment in the secondary phase must take account of emerging subject emphases at 11, and reflect a very largely subject-based organisation in the pre-14 stage and later up to 16 when it must articulate with GCSE and Records of Achievement (RoA). Profile components will need to draw on a wide and complex range of subject-based knowledge, skills and understanding, though it is desirable to include some general competences in profile components that all subject working groups agree to use in common.

159. While the national curriculum will render decisions about option choice at 14 less crucial, that age still represents an important staging-point because GCSE and RoA offer at 16 a mainly summative form of assessment - given added force, in the case of GCSE, by being a publicly recognised qualification (for many at the end of formal education).

Assessment at age 16

160. At 16, as we have already indicated, assessment related to the national curriculum will need to articulate with GCSE and other qualifications at

equivalent level, as well as with RoA. These are an integral part of the national curriculum provisions. Both in their intentions and in form, they underpin a cumulative picture of achievement built up (with interim reporting) on the basis of assessment confirmed and refined over time, and of work done in a variety of contexts.

161. Together the three initiatives offer, we believe, the prospect of more confident and wide-ranging assessment at 16. To the extent that they have a common purpose and focus, they offer the opportunity for a simpler, coherent and complementary approach to assessing and recording progress having a number of advantages: obviating the current duplication of time and work spent on separate aspects of assessment; keeping assessing, recording and reporting within manageable bounds; alleviating demands on teachers; and providing a focus for teacher development, strengthening their role as assessors and relating curriculum and assessment.

Records of achievement

162. In particular the moderation and national testing required by the national curriculum could furnish a basis for accreditation and validation for some components of the records of achievement. These in turn provide a means of recording achievement through the secondary years (indeed starting at the end of the primary phase), including that related to personal and social development; and all the more valuable if pupils are drawn formally into discussion with teachers and parents about their progress.

We therefore recommend the use of Records of Achievement as a vehicle for recording progress and achievement within the national assessment system.

An articulated system

163. The technical problem of the articulation of the national assessment levels and the GCSE grades might be solved by adopting numbered national assessment levels (1 to 10 as proposed earlier in this report) across the 7 to 14 age range, and retaining the lettered GCSE grading (A to G). For subjects not being

taken at GCSE, the national assessment levels 7, 8, 9 and 10 would be made available within profile components, and the other, lower levels used at age 14 might also be needed. However, the rules of combination of levels in profile components into GCSE subject grades are not our direct concern at this stage and we do not presume to specify how the link between national assessment profile components and GCSE grades is to be made. It will be sufficient to specify that at earlier reporting ages levels 6 and 7 signify that pupils are well on course for GCSE grades at age 16. It is not to be expected that attainment targets as elaborated in relation to the national curriculum, and the national/grade criteria attached to GCSE, will "meet" at all points.

We take it as axiomatic that eventually changes will be necessary to the GCSE and other criteria and that changes derived from the development of the national curriculum should have priority in an orderly process of amendment.

Summary of secondary school procedures

164. In summary, national assessment in secondary schools might work as follows (a fuller description is offered in Appendix M).

165. All secondary schools would receive on transfer detailed information on their pupils on the basis of assessment at age 11 against known attainment targets and programmes of study.

166. During the year in which the pupils become 14, teachers would be asked:

- on the basis of continuous assessment of work done (against programmes of study and attainment targets) and of inter-class moderation, to arrive at judgements of their pupils across a number of subject profile components;
- to administer nationally prescribed tests in the foundation subjects, to mark and then to analyse the results according to prescribed procedures;

- to attend a local group moderation meeting at which the teachers' judgements and the test results would be considered together and, with the help of samples and information on local and national distributions, to arrive at an overall assessment, and to resolve discrepancies;
- to compile an overall report related to 11 year old results for consideration by heads and governors for incorporation in a published report on the school related to its context;
- to adjust any individual pupil's result in the light of the moderation exercise, and then report on each pupil's performance for the pupil and parents concerned;
- to record results in a pupil's record of achievement;

167. For pupils aged 16, the same procedures would apply but only in respect of the subjects not being taken for GCSE.

XV THE ASSESSMENT OF CHILDREN WITH SPECIAL EDUCATIONAL NEEDS

168. Children with special educational needs make up a very diverse group. Indeed, it could be said that all children, at some time in their educational career, exhibit special teaching needs which have to be met. Children's learning difficulties may stem from physical and sensory disabilities, behavioural and medical problems, reading and communication difficulties or different degrees of mental handicap. They may be educated in special schools, special classes or units in ordinary schools, or may be integrated into ordinary classes. They may or may not be statemented: policies with regard to . statementing differ from one LEA to another. Some children with physical disabilities may nevertheless be scholastically extremely able and it is likely that, in their case, parents would rightly press for their inclusion in the national assessment programme.

Need for flexibility

169. Given the extremely diverse nature of special educational needs, it does not seem to us that it would be useful to issue blanket statements about the inclusion or otherwise of such children in the national assessment programme, but rather that the programme should have built into it sufficient flexibility to cope with a wide diversity of needs.

Like all children, those with special educational needs require attainable targets to encourage their development and promote their self-esteem. We therefore recommend that, wherever children with special educational needs are capable of undertaking the national tests, they should be encouraged to do so.

We further recommend that a special unit within the chosen test development agency should be dedicated to producing test materials and devising testing and assessment procedures sufficiently wide-ranging and sensitive to respond to the needs of these children.

Diagnostic materials

170. As for the school population as a whole, but even more crucially in this instance, we see the need for the provision of a wide range of diagnostic tests, to be made available to schools for use on an individual basis. These tests would need to be appropriate to children of a variety of ages, to be administered when they reached a particular level of development and attainment. Teachers of children with special educational needs should be able to draw on a wide range of diagnostic materials at any stage in the children's development, to help to determine the next steps to be taken to guide their personal and educational progress.

Coverage of ability range

171. The national tests themselves should be designed so as to be appropriate to children across the whole ability range, modified as necessary for children with particular sensory problems (ie blind and partially-sighted children, deaf and partially-hearing children) and those with communication difficulties. Strategies for ensuring that these requirements were met would need to be devised by the test development agencies during the pilot-testing phase of the test development process, but might include, for example, locator test items to direct children to suitable levels of difficulty, or the provision of items which would allow for a differentiated response. The recent development of computerised adaptive tests might well have an application here.

Exemption

172. Exemption from the national testing programme should, we think, be at the discretion of the head teacher if, in his or her professional judgement, testing would be likely to produce such stress that a medical or emotional condition would be exacerbated. The number of children so exempted should be a matter of record, to be justified during the moderation process. Children exempted from taking the national tests would provisionally be reported as so exempted. Some of them may be identified as in need of further specific help.

173. Some children with special educational needs might be entered for the national tests but might not prove capable of coping with them adequately. To avoid any possibility of such children perceiving themselves as 'failures', the national tests should be designed in such a way as to permit the teacher to curtail the test discreetly without the child being aware that this was being done, or to give assistance, the extent of this being recorded. (Experience with the Assessment of Performance Unit's Practical Mathematics tests has shown that this can be done sensitively and unobtrusively.)

IMPLEMENTATION

XVI THE BRIEF FOR THE SUBJECT GROUPS

174. This section draws together the proposals made in preceding sections, provides a review and summary of them, and spells out their immediate implications, in relation to the work of the national curriculum subject working groups.

Framework and exceptions

175. We propose that all the working groups set up for foundation subjects should be expected to operate within the common framework outlined in this report, but that exceptions from one or more of its requirements be allowed if substantial reasons for the exception can be provided. Such exceptions might be permitted more readily for non-core than for core subjects.

Development of profile components

176. Each subject working group should start from its list of attainment targets and combine them to form profile components as described in Section VI above. For level 2, some of the profile components will be broad, becoming differentiated as there is movement through the levels to level 10. Some early profile components might feed into a number of later subject components. In the later levels, the components should relate to such profile as might be formulated in the development of GCSE grade criteria.

177. Attainment for each profile component should be described in terms of a set of staged criteria, in relationship to which the diverse performances of children can be classified. The levels and corresponding criteria should form a single spectrum irrespective of age, but be so constructed that normal performance of pupils would correspond to the second level at age 7 (for profile components appropriate at that age) and would involve the use of a larger number up to 10 at later ages. These would span the complete range of GCSE grades, as specified in Section XI.

Relations with other subjects

178. It is possible, indeed desirable, that one or more of a subject's profile components might overlap strongly with, or even be identical with, components from another subject. Profile components from other subject groups should be studied with this in mind. Cross-curricular themes which are important in teaching should be reflected in overlap, or preferably identity, of appropriate profile components. Special attention should be paid to the need to limit the number of profile components at ages 7 and 11 by use of common components across subjects.

Each subject working group should decide on a limited number, usually 4, of profile components in relation to which any pupil's performance will be assessed and discussed. A criterion-referenced set of levels should be set out for each component, to span the full range of performance over the ages for which the component is applicable.

Assessment tasks and marking regimes

179. For each profile component, some tests or standard assessment tasks will have to be used, and the brief that a subject working group would want given to any agency responsible for construction and trial of these should be described, including any aspects of the marking which would be particularly significant.

180. Teachers' assessments will also form part of the assessment system. Only broad advice about these is needed, to guide teachers, although details on aspects likely to be unfamiliar might be given so that a case for particular assistance at national level might be considered.

181. For each profile component the methods for combination of standard and internal assessments should be as set out in Section X. If there are any particular problems with the recommended strategy for differentiation - by outcome at younger ages, by task where necessary at older ages - these should be explained.

Moderation

182. We have recommended that the results for each profile component be moderated by the procedures discussed in Section X. If it is felt that group moderation is really essential for some components, but dispensable for others, this should be explained.

183. Recommendations for age 16 will need to relate to the criteria in the appropriate GCSE subjects.

Subject working groups should specify, in broad terms and for each profile component, the appropriate tests (standardised assessment tasks) which should be prepared, and the advice and help which should be given to teachers about their corresponding internal assessments.

Aggregation for reporting

184. Whilst the levels within any one profile component are to be criterion-referenced, it will also be necessary, for reporting to some audiences, to aggregate across profile components to obtain a description of performance in the subject as a whole. The rules for such aggregation will have to be set out: how it may be carried out, whether by simple addition, or by weighted addition, perhaps involving some requirement that minimum levels of achievement must be attained in certain components or by rules which specify the single overall level secured by every one of the possible combinations of profile component levels.

185. The criteria for levels of the separate profile components may have to be related across components if such aggregations are to be sensible. This could be done by reference to national levels of achievement across profile components.

186. Care will have to be taken in determining ways in which results are reported to various users. Some attention should be given to the ways in which the various profile components and the levels within them can be clearly

described to the various audiences - pupils, parents, employers, governors and local and national government.

Combination of profile component levels to give a subject level should be by a specified procedure. Uniform ways of describing profile components and the level within each should be specified in language that is helpful to pupils, teachers, parents, employers and other users.

Relations with other agencies

187. It is not possible for us to make precise the boundary between the remit of subject working groups and that of those agencies, including particularly the new National Curriculum and the School Examination and Assessment Councils, who will have to commission construction of standard assessments and oversee operation of the new system. In particular, we note that the specification of resource implications following from the assessment recommendations of subject working groups is not explicitly in their remit.

Novelty and familiarity

188. Nevertheless, the subject working groups will be well equipped to give general advice on two aspects of need. The first is about the degree of novelty of the assessment tasks that need to be constructed: they will know whether what they are recommending for use are well-known and long-established methods, or innovations which will need extensive preliminary trial. The second is about the degree of familiarity, and therefore of confidence and capability, that teachers may be expected to enjoy in relation to the methods and procedures proposed. Again, general guidance here will be of help in the planning for INSET work that will have to be undertaken.

Subject working groups should give general advice about the degree of novelty of the assessments they envisage, so that the construction of them and the provision of in-service support for teachers can be appraised.

XVII PHASING

189. It is not possible on assessment grounds alone to set out a firm timetable for putting into operation the scheme that we propose. The assessment and testing is to be constructed in relation to the attainment targets and programmes of study in each subject, and the first determinant of the timetable will be the time required for these targets and programmes to be decided, and for schools to adapt to them.

Secondary schools

190. The situation in the secondary phase is more straightforward than in the primary because it is possible to talk of subjects separately. The following sequences may be envisaged for the work at age 14 and for work for pupils not taking GCSE at 16 (who will generally use the upper level materials for age 14:-

<u>Time</u>	<u>Activity in Secondary Range</u>
Start of Year 1	Subject Targets Promulgated.
During Year 1	Curriculum. Dissemination. Assessment Materials (at national level) in preparation for age 14.
Start of Year 2	Curriculum set in operation ages 11/12 and 12/13.
During Year 2	Further Development and Trials at age 14. Assessment.
Start of Year 3	Curriculum continues in operation 12/13, 13/14 and starts with new 11/12 group.
During Year 3	Teachers trained, and use assessment methods on trial basis. Some materials trialled for use with non GCSE pupils at age 16.
Start of Year 4	Curriculum in operation years 1-4 (Age 14/15).
During Year 4	First operational use of age 14 assessment with reporting.

Start of Year 5	Curriculum in operation years 1-5.
During Year 5	First operational use of the same materials at age 16 for non-GCSE pupils.

191. This is not a generous timetable. Given the unfamiliarity of the procedures and the need to adapt teaching methods to incorporate them, a trial run year is strongly recommended, at least at age 14. The timetable means that those following the system in the first year of formal operation at age 14 will have been taught within the national curriculum since age 11, although without the primary school background that the national curriculum will envisage. Those first assessed at 16 will have been taught within the national curriculum since age 12 and tested on a trial basis at age 14.

192. It will be important to stagger the implementation of the age 14 and of the age 16 work because the burden will usually fall on the same teachers. If 14 comes first, it becomes more easy to ensure that the first cohort to be involved at age 16 will have followed the national curriculum for 4 years.

GCSE

193. The above takes no account of GCSE developments. For at least the first four years, existing GCSE syllabuses and assessment methods should remain in operation. During that time, any changes in GCSE which follow from the definition of attainment targets up to age 16 in the national curriculum will have to be planned. An outline timetable might be:-

- Year 1 Agreed subject targets ready for consideration. New GCSE criteria prepared.
- Year 2 Examining groups prepare new syllabuses and specimen examinations.
- Year 3 New syllabuses and examinations approved and published.

Year 4 Schools start teaching to new syllabus for Year 4 GCSE pupils.

Year 5 First entry to new GCSE examinations.

194. Ideally revisions of GCSE assessment methods should arise from the evolution of the national system. However, a radical change in GCSE seems hardly feasible on the above timetable, which is only just possible on the assumption that the attainment targets and programmes of study for age 16, promulgated after consultation on the recommendations of the subject working groups, mean that the processes of consultation to settle new GCSE criteria can take a much shorter time than in the past. Even given this, the timetable would allow very little time for consultation, or for referral back of particular proposals of the group syllabuses stage. It may therefore be necessary to revise GCSE in two stages, one to incorporate the national curriculum targets within present systems of assessment and grading, followed later by more radical changes to the latter when the rest of the national assessment system had been in full operation for at least a short time. All of the above could be considered as timetables for an individual subject. If the year of commencement (year 1 in the tables) is not the same for all subjects, as seems almost certain, the adaptation in schools will be spread over a longer time.

Primary schools

195. For the primary phase, there is an additional complication that the separate subject proposals ought to be combined, so that primary teachers are working to an appropriate set of profile components as envisaged in Section XI. It seems inevitable that an interim set of profile components should be recommended for 7 and 11 when outcomes from enough subjects are available. Initial curriculum and assessment work would start with these and be adapted subsequently. It should be stressed that in the primary phase, in contrast to the secondary, the addition of guidelines deriving from further subject working groups may imply extra adaptation for the teacher who has already been operating with the outcomes of earlier working groups, rather than mere addition of extra profile components from each group. This could lead to an intolerable load on the primary teacher.

196. Given the above reservation, which applies to the fixing of year 1, the following timetable should be feasible:-

<u>Time</u>	<u>Activity in Primary Range</u>
Start of Year 1	Interim profile component Targets promulgated for age 7.
During Year 1	Curriculum. Dissemination. National assessment tasks for age 7 prepared.
Start of Year 2	Curriculum in operation ages 5/6. Interim targets for age 11 promulgated.
During Year 2	Trials of assessment tasks, Training starts age 7: tasks for age 11 prepared.
Start of Year 3	Curriculum in operation for ages 6/7. Also started for ages 8/9.
During Year 3	Teachers trained and use age 7 assessments on trial basis. Trials of age 11 materials.
Start of Year 4	Curriculum for all up to the 9/10 age group.
During Year 4	First operational use of age 7 assessment. Trials of age 11 materials and teachers trained.
Start of Year 5	Curriculum for all up to year 5 (10/11).
During Year 5	First operational use of age 11 assessment.

197. This has features similar to those for the secondary phasing. The staggering of ages 7 and 11 spreads out each school's adaptation load and leaves more time for any one group of children to be taught to the new targets over several years before the assessment system affects them. It also makes it possible to delay to year 2 the decisions about age 11 profile components, so that collaboration between more of the subject working groups can be arranged.

198. The above proposition accepts that pupils will reach the assessment ages with fewer years of experience of work adapted to the national curriculum attainment targets than they should have. For example, all of the first 5 years of secondary work and all of the first years of operation of age 14 and 16 assessments will be on those unaffected by the primary changes. Ideally, a new scheme would be implemented with an age 5 entry cohort and take 11 years to implement as they advance to age 16, but we recognise that the delay this would involve might be unacceptable for achieving national improvements in the work of secondary schools.

199. The times we have given are minimum times. If the implementation were to be more rapid, the work would start with assessment materials of mediocre quality, with pupils who have not had adequate time with the new curriculum, and with teachers inadequately prepared and feeling de-skilled by having to use procedures which they have not been able to assimilate into their own teaching. All of these features would make for a disastrous start and would set precedents from which the new system would take many years to recover. The phasing should pay regard to the undue stress on teachers, on their schools and the consequent harm to pupils which will arise if too many novel requirements are initiated over a short period of time.

We recommend that the new assessment system should be phased in over a period adequate for the preparation and trial of new assessment methods, for teacher preparation, and for pupils to benefit from extensive experience of the new curriculum. We estimate that this period needs to be at least 5 years from the promulgation of the relevant attainment targets.

A SYSTEM OF SUPPORT

XVIII THE SUPPORT STRUCTURE

200. The system of national assessment proposed in this report will call for the involvement of a number of agencies.

201. At the heart of the system will be the new School Examinations and Assessment Council and the National Curriculum Council, to co-ordinate the development of the national curriculum and the programme of assessment. It will also be essential for these Councils to arrange to provide advice on assessment to the various subject working groups, to ensure they furnish sufficiently clear statements of attainment targets and profile components to allow appropriate means of assessment to be developed. There would also need to be close liaison between the subject working groups and the chosen test development agencies, probably mediated through the two Councils.

Test development

202. The construction of the national tests would need to be contracted out to appropriate test development agencies. These agencies would have two main functions. The first would be the construction and trial of the national tests and assessment tasks to be used by all. The second would be the development of a range of supplementary tests and assessment procedures to support the educational process generally and to be used at the discretion of schools. The diagnostic tests would be closely focussed, applicable at a variety of ages, and would concentrate on analysis of methods and strategies used by pupils in order to identify their problems precisely and so to indicate how they should be helped. For both of these two functions, a range of agencies might be invited to take part. It is possible that some of the development for the second function could be undertaken by commercial test publishers as part of their normal programme.

Delivery

203. The delivery of national assessment and testing items and arrangements, the organisation of moderation procedures, and the collection and collation of results form a complex and interlinked set of functions. Maintenance of the resources, particularly of the collections of tests, whether broad standard assessment tasks or more sharply focussed types of test, will probably have to

be a central function. Distribution would include contacting schools, with an up-to-date register of all of them, and a service to deal with enquiries and to help with advice on training in marking and scaling procedures, and help in data analysis. This could be done either nationally by a central agency, or regionally through examining groups and/or LEAs. The moderation work would require organisation at a regional or local level and would have to be linked to INSET programmes for training teachers in the new assessment methods so that they could play their central role. The collection and collation of results might seem to be a routine matter of data analysis, but the outcomes of moderation work would be a valuable source of feedback about the quality of the national assessments, and such information ought to be carefully collected and analysed. Organisations capable of operation at regional and local level would be needed, and the LEAs or the examining groups, or both in collaboration, might be well equipped to undertake these functions.

Publications

204. To support those operating this system at all levels there will be a need for suitable publications, which the new Councils may have to commission and distribute.

Monitoring

205. The operation of the system ought to be carefully monitored. In particular, the effects of the new assessment and testing procedures on classroom teaching, on the overall curriculum and on existing assessment practices on schools, ought to be a subject for careful review, through commissioned research. A strategy for evaluation of the system and for its improvement, using a variety of sources of feedback, will have to be formulated, probably by the School Examinations and Assessment Council.

XIX INSET AND OTHER RESOURCE NEEDS

206. The assessment and testing upon which we have to advise is part of a broader programme of educational change. The introduction of a national

curriculum with its attendant attainment targets would, even if there were no implications for assessment, require additions to the facilities and resources of schools, the recruitment of teachers in appropriate distributions across schools and across the subject areas of the curriculum, and the pre-service and in-service education and training of these teachers to ensure that they were able to work well within the revised system. Therefore the needs associated with the changes in assessment have to be placed within this broader context.

In-service training needs

207. It will in particular be necessary for government and LEAs to ensure that appropriate in-service opportunities are put in place. Such opportunities will need to be designed so as to ensure that, first, all teachers understand the new assessment arrangements which support the national curriculum; and, second, teachers who are specifically involved in assessment at particular ages and/or in particular subject areas at both primary and secondary level are equipped with the necessary range of skills and techniques.

208. With the introduction of the GCSE, secondary schools and LEAs have had recent experience of the in-service programmes which of necessity accompany examination reform. The new arrangements for assessment cover the age range of 5-16 whereas the GCSE was essentially limited to the 14-16 age group; but in other respects the two developments have much in common. In both cases a revised view of the curriculum is being developed on the basis of good practice in the recent past, and the curriculum is to be associated with and supported by a change in the system of assessment. The INSET programme for the GCSE has been scrutinised by the DES and HMI, the SEC and the GCSE Examining Groups, and by the local authority associations. There is, therefore, a wide body of experience on which it would be wise to draw in formulating the larger INSET programme which revised roles for groups of teachers at four ages will inevitably require.

Facing new objectives

209. We draw special attention to only a few issues. It seems that in the GCSE more teachers have encountered problems in carrying out their new functions in

assessment than in accommodating to a revised syllabus or to a shift in the definition of the field of study. The use in the GCSE of a training programme which started with the national criteria and only in the later stages of training got down to the detail of individual syllabuses and their assessment systems meant that courses had often started before teachers had been able to think about the immediate concrete problems to be faced in their own schools, using a particular examining body. If, as in the GCSE, there is a shift in the balance of assessment objectives so that skills and abilities are given increased importance, this change alone requires more assessment time, more skill in the assessors and, in many cases, a reduction in the pupil/teacher ratio.

Materials needed

210. From these and similar points, we conclude that it would be unwise to start an INSET programme relevant to a particular assessment system until a range of assessment materials, of all types for the specified age group, could be brought into use in the training course. The types of assessment discussed in Section VII should all be brought under scrutiny and used in exercises in the course. Some of the practices and procedures use in "agreement trials" within the GCSE system could probably be modified to suit this INSET work.

Scale of operation

211. Apart from modifications which might arise for GCSE, the new work required for age 14 would imply a training operation roughly comparable in scale to the GCSE programme, perhaps greater because it would bear on secondary work up to age 14 which has not hitherto worked to constraints of any external system. For the primary phase, the whole operation would be more extensive than the GCSE programme because all the teachers involved will need to be trained in the assessment of all curricular areas. These needs should be carefully borne in mind in considering the phasing of introduction of any programme.

Other training needs

212. Other training needs will also arise. LEA staff of various kinds will need to be thoroughly conversant with the new system as part of their responsibilities in relation to the national curriculum. The management of the whole programme of assessment and the reporting thereof also points to the need for someone in each school (or within a group of small schools) to have the responsibility and skills for such work. The institutions which provide initial and in-service teacher training will need to be equipped to help with the new developments.

Other resource needs

213. In addition to in-service needs, the introduction of new arrangements will mean that a good deal of new work has to be undertaken if appropriate instruments for assessment and testing at the four age levels are to be available. Beyond the developmental stage there will be costs for the production, distribution and administration of sets of national test materials and of supplementary diagnostic items. At a later stage there will also be costs associated with whatever "reporting arrangements" are determined. It has not been practicable at this juncture in our work to quantify these.

214. We recognise that our proposals will require a substantial investment of time (particularly teacher time) and other resources. The level of investment will be conditioned by whatever decisions are taken on the final shape of the new system and the rate and phasing of its implementation. Detailed analysis will have to be undertaken; in the meantime we identify in Appendix N data about existing and planned provisions which should contribute to consideration of the implementation of a new system.

XX THE LOCAL EDUCATION AUTHORITY'S ROLE

215. We envisage that local education authorities will have a major part to play in implementing the proposals contained in this report. Essentially they will be concerned with the local delivery of the new arrangements and for meeting within their areas the requirements of the nationally determined framework.

They will be called upon to play a significant part in the preparation that will be necessary before a new system can be introduced. They will carry substantial responsibility for ensuring the implementation of the new arrangements at school level. They will be engaged in monitoring and evaluating the operation of the new system and in ensuring compliance with specific obligations (eg to establish a mechanism for dealing with complaints; to report on and publish information about the delivery of the national curriculum, including assessments).

216. We see LEAs as having responsibility for preparing and overseeing local strategies for giving advice and help to schools in relation to overall policies, and for training and supporting teachers largely through the establishment of specific INSET arrangements to meet the new needs. LEAs will need to work closely with other agencies involved in the delivery of a new assessment system. Decisions on the allocation and deployment of many of the resources which will be needed if the system is to be implemented in a planned and cost effective manner will rest with them. They may have a particularly important part to play, perhaps in co-operation with examining groups, in the arrangements for moderation using group meetings, as therefore both assessment and INSET functions. They will also be well placed to ensure that information about the new assessment arrangements and their outcomes is made available to the range of interested parties - schools will obviously be directly involved with parents, teachers and governing bodies, but on a wider front the general public and employers will need to be able to look to the LEA for this. In undertaking this task, one of the particular points to which LEAs will need to give attention is how best to set published information in the context of other relevant information about the service in their areas.

217. We believe that the existing strengths of LEAs will provide a good basis on which to build their contribution to the new system. They have at their disposal considerable experience of and expertise in curriculum matters generally and to some extent also in assessment/testing/examining arrangements. Some LEAs have already shown the value of co-operation between themselves and the examining boards and groups in the development and implementation of new forms of assessment. They have administrative and advisory support services available; they have mechanisms for ensuring the supply of books, materials,

accommodation and non-teaching support. These existing services provide a structure which can be used for the new task. They will however need to be enhanced. But the extent to which this is necessary will clearly depend on the degree of sophistication contained in the new arrangements, and cannot be considered in isolation from the general range of demands which the totality of the new national curriculum will undoubtedly place on LEA inspectorate/advisory services and on teachers in particular.

XXI FUTURE WORK FOR THE TASK GROUP

218. The problems and prospects considered in the last few sections of our report will require further and more detailed examination than we have been able to give so far. We propose to continue work on them and to produce a supplementary report. For this reason we make no substantive recommendations on these issues at this stage.

SUMMARY AND CONCLUSIONS

XXII CONCLUSIONS

219. Whilst the system we propose draws on many aspects of good practice that are already established, it is radically new in the articulation and comprehensive deployment of methods based on such experience. We are confident that the system we describe is practicable and can bring benefits to work both within schools and outside them. In particular, we can see how provision of new types of support within a framework of a new set of procedures can replace much of the large volume of testing and assessment at present in use. A co-ordinated system will use resources to better effect and will complement and support the existing assessment work that teachers already carry out. Thus the system should contribute to the raising of educational standards so that the broad educational needs of individuals and the national need to enhance the resources and skills of young people can be met.

Building on good classroom practice

220. As we stressed in our Introduction, the proposed procedures of assessment and testing bear directly upon the classroom practices of teachers. A system which was not closely linked to normal classroom assessments and which did not demand the professional skills and commitment of teachers might be less expensive and simpler to implement, but would be indefensible in that it could set in opposition the processes of learning, teaching and assessment.

Formative assessment to support learning

221. Our terms of reference stress that the assessment to be proposed must be "supportive of learning in schools". We reiterate that the four criteria set out in Section 1 are essential if this support is to be secured and we believe that they necessarily follow from the aims expressed in the consultative document on the national curriculum. The **formative** aspect follows almost by definition. For the system to be formative it must produce a full and well-articulated picture of the individual child's current strengths and future needs. No simple label 1-6 will achieve this function, nor is any entirely external testing system capable of producing the necessary richness of information without placing an insuperable load of formal assessment on the child. The formative aspect calls for profile reporting and the exercise of the professional judgement of teachers.

Raising standards

222. The system is also required to be formative at the national level, to play an active part in raising standards of attainment. **Criterion-referencing** inevitably follows. Norm-referenced approaches conceal changes in national standards. Whatever the average child accomplishes is the norm and if the average child's performance changes the reported norm remains the same figure. Only by criterion-referencing can standards be monitored. Only by criterion-referencing can they be communicated. Formative assessment requires the involvement of the professional judgement of teachers. Criterion-referencing helps to inform these judgements. Group **moderation** will enable the

dissemination of a shared language for discussing attainment at all levels - the central function of assessment. These three features will help to emphasise growth. They result in **progression** - a key element in ensuring that pupils and parents receive focused and evolving guidance throughout their school careers. Consistent and de-motivating confirmation of everything as it was at the previous reporting age can be avoided only if pupils and parents can have clear evidence of progress by use of the single sequence of levels across all ages in the way that we propose.

The unity of our proposals

223. We have considered systems of assessment and testing which are very different from the one that we propose. All alternatives impoverish the relationships between assessment and learning, so that the former harms the latter instead of supporting it. Most of them give no clear information or guidance about pupils' achievement or progress, and they all risk interference with, rather than support of, teachers' work with pupils. Thus we cannot recommend any simpler alternative to our proposals. There is of course room for variation in their implementation: for example, using group moderation procedures for a restricted number of profile components, or not using all such procedures on every annual assessment cycle, or phasing in more slowly to spread the load on teachers and on the support systems. None of these marginal changes would destroy essential features of our system, although they might weaken its impact in the short term. However, any major change that we can envisage would destroy the linked unity of our proposals and lose most of the benefits which they are aimed to secure within and for the national curriculum.

Securing teachers' commitment

224. The underlying unity of the three aspects of education - teaching, learning, and assessment - is fundamental to the strategy which informs our proposals. The strategy will fail if teachers do not come to have confidence in, and commitment to, the new system as a positive part of their teaching. Securing this commitment is the essential pre-condition for the new system to realise the considerable value that it could bring.

225. Amongst the conditions which will have to be met to secure this professional commitment will be the following:-

- Clear acceptance that the aim is to support and enhance the professional skills that teachers already deploy to promote learning.
- Clear recognition that the focus of responsibility for operation of a new system lies with teachers within schools.
- Stress on the formative aims and on giving clear guidance about progress to pupils and to their parents.
- Widespread consultation and discussion before proposals are put into effect.
- A realistic time-scale for phasing in a new system.
- Adequate resources, including in-service provision.
- Help with moderation procedures so that the system contributes to communication within schools, between schools, parents and governors, and to the community as a whole about the realisation and evaluation of the aims of schools.
- Sensitive handling of any requirements for outside reporting, recognising that simplistic procedures could mislead parents, damage schools, and impair relations between teachers and their pupils.

226. If there is one main motive to explain our support for the system we propose, it is that we believe that it can provide the essential means for promoting the learning development of children: support for teachers in enhancing the resources and professional skills which they deploy.

XXIII. A LIST OF RECOMMENDATIONS

227. The recommendations we have made in this report are listed below.

PURPOSES AND PRINCIPLES

1. The basis of the national assessment system should be essentially formative, but designed also to indicate where there is need for more detail diagnostic assessment. At age 16, however, it should incorporate assessment with summative functions. (Paragraph 27)
2. All assessment information about an individual should be treated as confidential and thus confined to those who need to know in order to help that pupil. (Paragraph 28)
3. For summative and evaluative purposes results should be aggregated across classes or schools so that no individual performances can be separated out. (Paragraph 29)
4. Assessment of attitudes should not form a prescribed part of the national assessment system. (Paragraph 30)
5. To realise the formative purpose of the national assessment system, pupil results in a subject should be presented as an attainment profile. (Paragraph 33)
6. An individual subject should report a small number (preferably no more than four and never more than six) of profile components reflecting the variety of knowledge, skills and understanding to which the subject gives rise. Wherever possible, one or more components should have more general application across the curriculum: for these a single common specification should be adopted in each of the subjects concerned. (Paragraph 35)

7. The national system should employ tests for which a wide range of modes of presentation, operation and response should be used so that each may be valid in relation to the attainment targets assessed. These particular tests should be called "standard assessment tasks" and they should be so designed that flexibility of form and use is allowed wherever this can be consistent with national comparability of results. (Paragraph 50)
8. Assessment tasks should be reviewed regularly for evidence of bias, particularly in respect of gender and race. (Paragraph 52)
9. Attainment targets should be exemplified as far as possible using specimen tasks. Such tasks can then assist in the communication of these targets. (Paragraph 56)
10. A mixture of standardised assessment instruments including tests, practical tasks and observations should be used in the national assessment system in order to minimise curriculum distortion. (Paragraph 59)
11. Teachers' ratings of pupil performance should be used as a fundamental element of the national assessment system. Just as with the national tests or tasks, teachers' own ratings should be derived from a variety of methods of evoking and assessing pupils' responses. (Paragraph 60)
12. When the subject working groups provide guidance on the aggregation of targets into a small number of profile components, they should have regard to the need for each component to lead to a report in which reasonable confidence is possible. (Paragraph 61)
13. Teachers' ratings should be moderated in such a way as to convey and to inform national standards. (Paragraph 62)
14. The national assessment system should be based on a combination of moderated teachers' ratings and standardised assessment tasks. (Paragraph 63)
15. Group moderation should be an integral part of the national assessment system. It should be used to produce the agreed combination of moderated teachers' ratings and the results of the national tests. (Paragraph 77)

16. An item bank of further assessment instruments should be available for teachers to use in cases where they need additional evidence about particular pupils. (Paragraph 78).

17. The final reports on individual pupils to their parents should be the responsibility of the teacher, supported by standardised assessment tasks and group moderation. (Paragraph 80)

18. Wherever schools use national assessment results in reports for evaluative purposes, they should report the distribution of pupil achievements. (Paragraph 84)

THE ASSESSMENT SYSTEM IN PRACTICE

19. The ages for national assessment should be 7, 11, 14 and 16; with reporting occurring near the end of the school year in which each cohort reaches the age involved. (Paragraph 92)

20. Each of the subject working groups should define a sequence of levels in each of its profile components, related to broad criteria for progression in that component. For a profile component which applies over the full age-range 7 to 16, there should be ten such levels, with corresponding reduction for profile components which will apply over a smaller span of school years. (Paragraph 101)

21. Levels 1 to 3 should be used for national assessments at age 7. (Paragraph 103)

22. The formal relationship between national assessment and GCSE should be limited, in the first instance, to this one reference point: and accordingly the boundary between levels 6 and 7 should correspond to the grade F/G boundary for GCSE. (Paragraph 105)

23. As they develop the upper four levels of their profile components, the subject working groups should adopt present practices for determining GCSE grades at A/B, C/D, mid-E, and F/G as a starting point. (Paragraph 106)

24. GCSE should be retained in its present form until the national assessment system is initiated at earlier ages. (Paragraph 107)
25. Assessment and reporting for the national assessment system should be at the same ages for all pupils, and differentiation should be based on the use of the single sequence of levels set up to cover progression over the full age range. (Paragraph 112)
26. Support items, procedures and training should be provided to help teachers relate their own assessments to the targets and assessment criteria of the national curriculum. (Paragraph 116)
27. A review should be made of the materials available to schools for detailed diagnostic investigation of pupils' learning problems, and that the need for extra help with production or advice about such materials should be considered. (Paragraph 117)
28. A working group should be established, with some shared membership between the subject working groups, to co-ordinate their proposals for assessment, including testing, at the primary stages, in the light of a comprehensive view of the primary curriculum and of the need to limit the assessment burden on teachers. (Paragraph 123)
29. National assessment results for any individual pupil should be confidential, to be discussed between pupil, parents and teachers, and to be otherwise transmitted in confidence. National assessment results for a class as a whole and a school as a whole should be available to the parents of its pupils. (Paragraph 131)
30. The only form in which results of national assessment for, and identifying, a given school should be published should be as part of a broader report by that school of its work as a whole. (Paragraph 132)

31. Any report by a school which includes national assessment results should include a general report for the area, prepared by the local authority, to indicate the nature of socio-economic and other influences which are known to affect schools. This report should give a general indication of the known effects of such influences on performance. (Paragraph 134)
32. National assessment results, for pupils at age 11, aggregated at school level, should be published as part of each primary school's report. There should be no requirement to publish results for pupils at age 7. (Paragraph 137)
33. National assessment results for pupils at ages 14 and 16, aggregated at school level, should be published as part of each school's report. (Paragraph 138)
34. At age 7 the standard assessment tasks for the national assessment should comprise a choice of three prescribed tasks for each child; each task should be designed to give opportunities for systematic assessment of competence in the range of profile components appropriate to age 7.
35. At age 11 the tests for national assessment should include three or four standard tasks which cover a range of profile components, possibly supplemented by more narrowly focussed tests for particular components. (Paragraph 153)
36. Records of Achievement should be used as a vehicle for recording progress and achievement within the national assessment system. (Paragraph 162)
37. Eventually changes will be necessary to the GCSE and other criteria. Changes derived from the development of the national curriculum should have priority in an orderly process of amendment. (Paragraph 163)
38. Like all children, those with special educational needs require attainable targets to encourage their development and promote their self-esteem. Wherever children with special educational needs are capable of undertaking the national tests, they should be encouraged to do so. (Paragraph 169)

39. A special unit within a chosen test development agency should be dedicated to producing test materials and devising testing and assessment procedures sufficiently wide-ranging and sensitive to respond to the needs of these children. (Paragraph 169)

IMPLEMENTATION

40. Each subject working group should decide on a limited number, usually 4, of profile components in relation to which any pupil's performance will be assessed and discussed. A criterion-referenced set of levels should be set out for each component, to span the full range of performance over the ages for which the component is applicable. (Paragraph 178)

41. Subject working groups should specify, in broad terms and for each profile component, the appropriate tests (standardised assessment tasks) which should be prepared, and the advice and help which should be given to teachers about their corresponding internal assessments. (Paragraph 183)

42. Combination of profile component levels to give a subject level should be by a specified procedure. Uniform ways of describing profile components and the level within each should be specified in language that is helpful to pupils, teachers, parents, employers and other users. (Paragraph 186)

43. Subject working groups should give general advice about the degree of novelty of the assessments they envisage, so that the construction of them and the provision of in-service support for teachers can be appraised. (Paragraph 188)

44. The new assessment system should be phased in over a period adequate for the preparation and trial of new assessment methods, for teacher preparation, and for pupils to benefit from extensive experience of the new curriculum. This period needs to be at least 5 years from the promulgation of the relevant attainment targets. (Paragraph 199)

APPENDICES

- A. Terms of Reference and Membership
- B. Letter of Guidance from the Secretary of State
- C. Meetings with Invited Groups
- D. Test Tasks for 7 and 11 Year Olds
- E. Test and Assessment Methods: Examples of the Variety Available
- F. National Testing and Equal Opportunities: A Paper by the Equal Opportunities Commission
- G. Technical Note on Reliability, Validity and Error
- H. The Conduct of Group Moderation Meetings
- I. Records of Achievement
- J. Technical Note on Background Factors Affecting Educational Performance
- K. Numbers in Schools
- L. How the System Might Work in a Primary School
- M. How the System Might Work in a Secondary School
- N. Resource Needs: Some Data on Existing and Planned Provision

TASK GROUP ON ASSESSMENT AND TESTING

TERMS OF REFERENCE

1. To advise the Secretary of State on the practical considerations which should govern all assessment including testing of attainment at age (approximately) 7, 11, 14 and 16, within a national curriculum; including the marking scale or scales and kinds of assessment including testing to be used, the need to differentiate so that the assessment can promote learning across a range of abilities, the relative roles of informative and of diagnostic assessment, the uses to which the results of assessment should be put, the moderation requirements needed to secure credibility for assessments, and the publication and other services needed to support the system - with a view to securing assessment and testing arrangements which are simple to administer, understandable by all in and outside the education service, cost-effective, and supportive of learning in schools.
2. In making recommendations, to take into account the need not to increase calls on teachers' and pupils' time for activities which do not directly promote learning, and to limit costs.
3. To advise on the possible staging of the introduction of assessment including testing to reflect the need for the process to be manageable and for teachers to be adequately trained.
4. To report to the Secretary of State by Christmas 1987.

MEMBERSHIP

CHAIRMAN: Professor P J Black OBE, KSG, BSc, PhD FIP - Professor of Science Education at the University of London and Head of the Centre for Educational Studies at King's College, London (KQC).

MEMBERS: Professor J T Allanson - Emeritus Professor of Electronic and Electrical Engineering and member of the Secondary Examinations Council;

Mr J A Barnes - Director General of the City and Guilds of London Institute;

Dr C Burstall - Director of the National Foundation for Educational Research;

Mr O Hughes - Headteacher of a secondary school in Clwyd;

Mrs D Kavanagh - Headteacher of a primary school in Oxfordshire;

Mr J O Morris - County Education Officer, Essex County Council;

Mr D A T N Payne - Director of Personnel, Rolls Royce PLC;

Ms H Steedman - Researcher at the National Institute for Economic and Social Research, London;

Mr N Thomas - Visiting Professor of Education, Warwick University, honorary Professor of Education at the North East London Polytechnic and former Chief Inspector for Primary Education in HM Inspectorate of Schools.

CO-OPTED: Mr T Christie - member of the University of Manchester Department of Education, and the Northern Examining Association.

OBSERVER: Mr B C Arthur (HMI)

SECRETARIAT:

Ms G G Beauchamp

Mrs H J Field

Mrs L F Cooke

IN
ATTENDANCE
FROM THE
DEPARTMENT
OF EDUCATION
AND
SCIENCE:

Miss J H Bacon

Mr M D Phipps



ELIZABETH HOUSE
YORK ROAD
LONDON SE1 7PH
01-934 9000

Professor Paul Black OBE
King's College London
Chelsea Campus
Centre for Educational Studies
552 King's Road
LONDON SW10 0UA

30 October 1987

NATIONAL CURRICULUM: TASK GROUP ON ASSESSMENT AND TESTING

1. The purpose of this letter is to offer more detailed guidance on your task than is contained in your Terms of Reference.

Relationships with the Subject Working Groups

2. You will have seen the letters of guidance which I have sent to the Chairmen of the Mathematics and Science Working Groups. I have asked those Groups to recommend attainment targets which provide objectives specific enough to offer a sound basis for the assessment and testing of pupils' progress and performance; and programmes of study detailing the content, skills and processes which all pupils need to be taught. I have made clear that the main focus of those Groups' work is to be on attainment targets and programmes of study, but I have asked them, after consultation with your Group, to offer advice in broad terms by the end of November about assessment in relation to the attainment targets they are developing, and particularly what might appropriately be measured by nationally-prescribed tests. I have also asked them to offer general advice in their final reports (June 1988) about assessment, including testing, in relation to their subjects, which should take account of the overall strategy to which I hope your recommendations will lead.

3. I shall be sending broadly similar letters to other Working Groups for foundation subjects when they are set up. Your Group will need to liaise closely with the operational Subject Working Groups, and with the Committee on English chaired by Sir John Kingman. The process of developing attainment targets, programmes of study and an overall system of assessment, including testing, will be iterative.

Scope of the task

4. I seek advice on a coherent system of assessment, including testing, to cover the whole period of compulsory schooling. I hope that your Report will include general criteria applicable to the system as a whole but capable of accommodating appropriate variation eg between subjects; and more specific criteria for the choice and use of

assessment and testing arrangements for particular subject areas or age-ranges. The latter in particular should take account of discussion with the operational Subject Working Groups and with informed people and organisations in other subject areas.

5. The Group's recommendations must be practicable to implement and cost-effective. I hope that your recommendations will take account of the very considerable amount of assessment which is already carried out as a normal part of teaching and learning in our schools, and will recognise that all forms of assessment affect the teaching and learning assessed. I am looking for arrangements which, by supplementing the normal assessments made by teachers in the classroom with simply-administered tests, will offer a clear picture of how pupils, individually and collectively, are faring at each of the age points. Such arrangements should help to promote good teaching.

The purposes and nature of assessment

6. Our consultation document recognises a number of purposes which assessment, including testing, may fulfil, and different uses to which information derived from assessments may be put. These include diagnostic or formative purposes, mainly concerned with ascertaining what stage a pupil has reached, identifying strengths and weaknesses and planning the appropriate next steps in the pupil's education; summative purposes, concerned with recording in a systematic way the pupil's achievement overall, particularly compared with attainment targets for each subject; and purposes mainly concerned with publicising and evaluating the work of the education service and its various parts in the light of pupils' achievements. I attach importance to all of these, and expect your recommendations to cover all. But as I have said in my letter of guidance to the Chairmen of the Science and Maths Working Groups, the main purpose of the national assessment, including testing, in relation to agreed attainment targets for national curriculum foundation subjects will be to "show what a pupil has learnt and mastered, so as to enable teachers and parents to ensure that he or she is making adequate progress and to inform decisions about the next steps".

7. I seek advice on a coherent national system for assessment and testing in relation to agreed attainment targets that recognises the different purposes to which assessment is put and takes account of how they interrelate with and complement each other, bearing in mind the points that I set out in paragraph 5 above.

8. The Group will no doubt consider a range of assessment approaches including written, practical and oral tests, and a variety of arrangements for administering, marking and moderating assessments and reporting their results. I seek advice on how these approaches should come together within a commonly-applicable system to achieve the objectives I have spelt out above.

The amount of assessment

9. I hope that the Group's advice will cover both the overall volume and the combinations of assessment and testing methods which can reasonably be accommodated, either within the normal teaching process or alongside it, but without adverse effects upon teaching time or pupils' workloads.

Ages and stages

10. Our consultation document envisages four formal assessment points during compulsory schooling. There is already assessment at age 16 for most pupils through GCSE and other public examinations. My assumption is that GCSE, or equivalent level examinations approved against GCSE criteria, will provide the basis for assessment in the foundation subjects at 16+ for virtually all pupils, and that consequently the GCSE criteria will be revised as appropriate to reflect agreed attainment targets and programmes of study and to ensure that the examinations present appropriate challenges

to pupils at all levels of ability. The Group's recommendations will be relevant for those conducting this revision of the GCSE national criteria.

11. Our consultation document proposes that other assessment points should be at age 14 - when decisions about examination courses to be pursued at age 16 are normally taken; age 11 - the most usual transition point between primary and secondary education; and around the age of 7, the precise age for the first assessment stage for each subject depending on the advice of subject working groups. I expect the Group's advice to distinguish what forms of assessment, including testing, would be most appropriate for each of these age points. I assume that the reporting of the results of assessment, including nationally prescribed testing, should take place towards the end of the school year in which the majority of pupils in a teaching group reach one of the age points, although cumulative assessment contributing to the final grading may take place throughout the stage preceding that. If the Group finds compelling arguments for ages and stages other than those proposed in order to offer a coherent system, it should so advise me. I shall also particularly value advice on how assessment and testing, and interpretation of the results at the first age point (7 or thereabouts) can properly take account of the wide variation in pupils' starting points in formal education, including their previous experience in pre-school education.

Differentiation

12. I am asking the subject working groups to recommend attainment targets for the knowledge, skills and understanding which pupils of a range of different abilities should normally be expected to achieve at the four age points; but so far as possible to avoid setting qualitatively different targets - in terms of areas of knowledge, skill or understanding - for children of different ability. My objective is that teachers should not have to prejudge what attainment targets are appropriate for a particular pupil; differentiation should be by outcome not pre-selection. I therefore seek advice on the measurement and recording at a range of different levels of positive achievement in reasonably discrete elements, within each common target area. Some elements might constitute essential performance targets for all pupils; others might allow for a series of graduated outcomes.

13. I shall particularly welcome advice on the application of such an approach to the two extremes of the performance range - the very able, and the least able.

Reporting on the results of assessment

14. I attach considerable importance to improving communication and understanding at various levels about educational objectives and performance; and information derived from assessment and testing will play a key part in that. I seek advice on the information which will be required to meet the purposes identified by the Group, taking account of the need for a sufficiently detailed profile of a pupil's performance to give a full and fair indication of attainment, but also of the need to be able to present aggregated information in ways that will be useful to a wider public than the teacher, pupil and his or her parents. I should welcome advice on procedures for ensuring consistency and compatibility between different approaches to assessment, including testing, and for aggregating results. I also seek advice on a standard scale or scales of marks for describing attainment, and would welcome the Group's views on the potential role of records of achievement in recording and reporting performance and attainment, at all stages of schooling. The Group should advise on the practical implications of the reporting and recording arrangements it proposes for schools and LEAs.

Administration and moderation arrangements

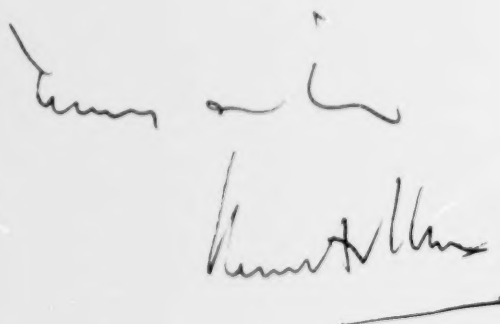
15. Our consultation document suggests that arrangements will need to be made

for the development, piloting and standardisation of assessment instruments including tests; for storing and retrieving such instruments; and for distributing them to schools. I should be grateful for any advice the Group wishes to offer on such arrangements in the light of the system it recommends.

16. The consultation document also proposes that assessment including the administering of nationally prescribed tests should be done in schools by teachers; but that arrangements should be made, probably with the GCSE examining groups, for external moderation of this work. Again I should be grateful for the Group's views on cost-effective and practical arrangements for securing "quality control" of school-based assessments, so as to secure credibility and confidence within the education service and with the public at large.

Conclusion

17. I look forward to receiving the Group's report by Christmas 1987. I recognise that it may not be possible to cover all the ground above by then, but hope that you will be able to advise on the main issues in paragraph 1 of your terms of reference and paragraphs 4-13 above, leaving other issues and fuller elaboration for later advice. I shall make your report, together with my conclusions, available for the guidance of the subject working groups, and intend that it should be published.

A handwritten signature in dark ink, appearing to read 'Kenneth Baker', with a horizontal line drawn underneath it.

MEETINGS WITH INVITED GROUPS

During the course of its work the Task Group on Assessment and Testing met with various individuals and organisations to collect evidence and discuss issues.

1. Graded Assessment

Representatives of THE JMB Staged Assessment in Literacy, London Graded Assessment in Mathematics, the Oxford Certificate of Educational Achievement, the Kent Mathematics Project, London Graded Assessment in CDT, and the London Graded Assessment in Science Project submitted materials and spoke about the work of their schemes. Discussion centred on the motivating effect of clear objectives and short term goals, credibility in terms of public recognition and confidence, and resource implications including development costs, INSET and moderation. Developing practice in graded assessment in modern languages was also considered (see Page and Hewitt "Languages Step by Step" CILT).

2. The Assessment of Performance Unit

Representatives of the Mathematics, Language, Foreign Languages, Science, and Design and Technology teams submitted materials and spoke about the work of their schemes. Discussion covered on specifying components within subject areas, the purposes of testing, the development and delivery of assessment tasks, the time required, the range of assessment tasks and INSET implications.

3. Chief Education Officers

Chief Education Officers representing the local education authorities of Hampshire, Lancashire, ACC, Trafford, AMA, Croydon, Northumberland, and ILEA, together with the Education Officers of the ACC and AMA submitted materials and spoke about assessment in their authorities. Discussion covered the extent to which assessment was employed, the purposes it served (whether summative, formative or diagnostic), the scope and nature of assessment and testing, resource implications and reporting results. Data on the extensive use of testing in local education authorities was also obtained, from the NFER records and from the study "Testing Children" by Gipps, Steadman, Blackstone and Stierer (Heinemann 1983).

4. Experts on and Practitioners of Assessment in Other Countries

Experts on assessment systems in other countries, together with Headteachers from Germany and France spoke about the work of their assessment schemes. The discussion centred on systems operating in France, the Federal Republic of Germany, the Netherlands, Sweden, Canada, and the USA. Issues discussed included the extent of the use made of external standardised tests, use of other tests (including diagnostic tests), the nature and scope of in-school assessment, the scale of marks or grades, the arrangements made for recording and reporting, and the relationship of assessment systems to external (qualificatory) examination systems.

5. The Kingman Committee

Members of the Kingman Committee spoke about their work. Issues discussed included the moderation of assessment and testing, the task-specific nature of performance in English, differentiation by outcome and by pre-selection of tasks, attainment targets, the assessment of children at the age of 7, and the need to avoid cultural, racial, linguistic and gender bias in assessment instruments.

6. School Assessment Systems

Head teachers of primary and secondary schools in ILEA, West Glamorgan, Cornwall, and Bedfordshire submitted materials and spoke about the work of their assessment schemes. Discussion highlighted the multiplicity of assessment arrangements within schools and between schools, the purposes of assessment, the uses to which results are put, the importance and value of moderation, particularly by teams of teachers, reporting to parents and resource implications.

7. Science and Mathematics Working Groups

Close contact was maintained throughout by joint meetings of the chairman with the chairmen of these two groups, and by his attendance at each group for particular discussion of TGAT's evolving policy.

This appendix gives informal accounts of two tasks actually carried out in primary schools. Although these were not developed and carried out as standard assessment tasks, they will serve to illustrate how carefully structured class activities could provide opportunities for systematic assessment, in a range of profile components, within a context of activity that was meaningful and motivating for a primary pupil. It is recognised that extensive development would be used before these could be adequate for national assessment purposes - here they serve to illustrate in general terms the type of activity envisaged.

FIRST EXAMPLE - A TASK FOR SEVEN-YEAR-OLDS

The following task was one of several used with an infant class of 5-7 year old children as part of a topic entitled "Winter through to Spring". The task involved estimation, observation, measuring, drawing sensible conclusions and recording, and there were opportunities at various points throughout for the assessment of profiles in Language (oral and written), Mathematics and Science (problem-solving and measurement). The children, in the main, worked in groups and arrived at conclusions collaboratively although there was ample opportunity for individual work.

First activity - Discussion, interviewing, collecting data

LANGUAGE	Question asked: How do we keep warm in winter? Discussion.
MATHS	Make a survey of the clothes worn by the children in the classroom.
MATHS/ART	Record this information (graph).

Second activity - Estimation, measurement of temperature and time.

LANGUAGE	Question asked: Which materials keep us warmest? Wrap one layer of a different material around each of 6 milk bottles - cotton, fur, wool, felt, newspaper, nylon.
----------	---

SCIENCE AND
MATHS

Fill the bottles with warm water.
Measure the temperature in each, using a thermometer, and record it.
Estimate which bottle of water will cool first.
Measure the temperature every half-hour and record.
Which bottle of water grows cold first?
Why?
Which stayed warm?
Why?

Third activity - Estimation, measurement of temperature and time, problem solving, recording, discussion, sequencing, writing, computer work.

SCIENCE

Question asked: What would help to keep the bottles of water warm longer?

LANGUAGE

Discussion.

SCIENCE

Repeat the experiment, first wrapping four layers of one type of material around each bottle.
Are the results different?
How?
Why?
Which material kept the water warm longest?

MATHS

Put the materials into sequence, coolest to warmest.

HEALTH EDUCATION

How is it best to dress in cold weather?

LANGUAGE

Write, or use the computer to write, the story of this experiment.

SECOND EXAMPLE - A TASK FOR ELEVEN-YEAR-OLDS

The following is a task that has been used in several primary schools by teachers on an INSET course designed to develop use of good materials for pupils to learn by active engagement. The activity was to involve both the hand and the brain. Discussion by pupils, in groups and as a class, was an important feature: practical tasks have the advantage of giving pupils something to talk about. The task is specified below, in outline only, as a sequence of eight activities. These were covered in about four afternoon sessions. The experience showed that the task is too difficult in this form for 7 year olds, but suitable for 11 year olds.

MATERIALS TASK

First activity - Discussion

What do the children know?

Question asked: What different materials are there in the room?

OBSERVATION
(SCIENCE)

i. Collecting children's ideas

LANGUAGE

ii. Reviewing children's words and distinguishing between 'material' words and 'object' words.

Second activity - Collecting and Sorting

Children bring in different materials and sorting them.

OBSERVATION
(SCIENCE)

Discussion: what criteria were used for sorting? What were the difficulties in sorting?

CLASSIFYING
(MATHEMATICS)
DISCUSSION/
LANGUAGE

"Guess the group" game.

Third activity - Representation of Sorting Activity

MATHEMATICS

What different groups could selected items belong to?
(Matrix diagrams: 1 x 2, 2 x 2 matrix).

SCIENCE

Children sorting by two attributes - to move them from the notion of a material to the idea of a "characteristic" of a material.

LANGUAGE

Fourth activity - Usefulness of Sorting

Why do people group things in particular ways?

LANGUAGE

Children describing and sorting different woods.

Discussion: Which people need to know about wood?

What do they need to know? (Focus on characteristics of wood.)

Fifth activity - Re-sorting Materials into Metals and Non-Metals

SCIENCE

Which people need to know about metals?

LANGUAGE

Discussion: children's ideas about metals - the words they use to describe them and the uses they attribute to them.

OBSERVATION
(SCIENCE)

Focus on one use - a clip. Children observing the use in pen clips, hair grips, slides and safety pins.

Discussion: children describing the job to be done by the metal and what the metal has to be like (focus on metallic properties).

Sixth activity - Which Metal is the Best Bender?

SCIENCE Children grappling with what to measure, whilst carrying out a trial and error investigation.

LANGUAGE Discussion: What do we mean by best bender?

Seventh activity - Discussion of Investigation

SCIENCE Children evaluating their trial and error investigations.

LANGUAGE Focussing on: purpose for the investigation, ie information
MATHEMATICS for pen manufacturer, type of information needed, and what has to be judged.

Eighth activity - Further Investigation

MATHEMATICS Focus on how to measure and controls.

SCIENCE

The labels on the left indicate where opportunities would arise for assessment of various skills. The way in which this could be done in standardised form is not set out here: substantial development and trialling would be needed to establish such methods. There are several opportunities for monitoring discussions, for observing practical work by pupils working in groups, and through written work. The task could clearly be broken into two parts (after the fourth activity) but its length was not a disadvantage in practice. It would not be difficult to arrange for each assessed component to be involved in at least two points over the eight activities, thus giving more time for a teacher to assess all pupils once on that component during the whole task.

Important lessons learnt from the use of this and similar tasks are as follows:

- the conceptual understanding of the subject of a task can be a hurdle for many children: if this is not overcome their other competences cannot be fully assessed.

- topics must be so chosen that they give a wide range of possible development - this allows differentiation to be developed within the task as children show competence or difficulty at early stages, and also ensures that a range of profile components can be assessed.
- the task is both curriculum/learning material and assessment material: teachers find it very hard at first to keep both aspects in balance.
- in particular there is a temptation to close down open discussions and over-define situations or outcomes, partly because of a desire to "teach" and partly because of a belief that the assessment context must be rigidly defined.
- the use of group work, encouraging discussion within and reporting by groups, and of genuine class discussion, exploring children's ideas and results and not inserting teachers' own too early, is an unusual style for many, who find it hard to adapt.
- teachers need more than initial instruction - they also need mutual support through meetings, en route during the work, to exchange experiences and give mutual comfort.

These tasks were not design to cover (what might become) the full range of profile components at the primary ages. For example, opportunities to use artistic or craft skills might be taken up if they were re-designed for this broader purpose.

TEST AND ASSESSMENT METHODS: EXAMPLES OF THE VARIETY AVAILABLE.

The intention behind this Appendix is to provide a selection of forms of assessment and types of test in order to illustrate the range and variety of testing techniques which are available or are capable of development. It should not be seen as a full compendium of tests and test types, but rather as a pot-pourri of possibilities. Two further examples are included in Appendix D: these are set apart because they illustrate a particular approach recommended for testing in the primary phase.

As used in this Appendix and elsewhere in the report, a test is a formal procedure for evaluating aspects of an individual's knowledge, skills and understanding - or 'behaviour'. It provides a means of eliciting samples of behaviour in a standardised manner and prescribed procedures for grading it. The breadth of this definition indicates the great variety of approaches which are encompassed by tests. They differ in their modes of presentation, their modes of working and in the modes of responding. Modes of presentation can include oral questions, written questions, video films, computer displays and demonstrations. Modes of working can be mental, written, oral, practical and combinations of these. The responses, or products, of the test can be in written or oral form or require active performance, constructions or combination of these forms.

Tests differ also in the size of the samples of behaviour they assess. Some tests are combinations of single short questions; others have tasks which require the integration of several skills and a variety of aspects of knowledge. Both these approaches have advantages and disadvantages which make them more or less appropriate in specific situations.

Some of the contrasts between tests composed of discrete questions and those consisting of integrated tasks are shown in Table 1. Essentially, tests consisting of a series of short questions can sample more widely and have greater control over standardisation; integrated tasks provide more realistic samples of behaviour and may therefore provide more valid measures of attainment

particularly for younger pupils. In the integrated tasks, the emphasis on formality is reduced and, for example, interactions with assessors may be included some of it as part of the assessors work of gathering results. This interaction is, though, still subject to control and a set of rules would still govern the information imparted in the exchanges.

The choice of a particular test or testing technique depends on several factors. Primary among these is the purpose for which the results will be used but practicalities can also be important. These would include the time available, the resources available, the age and other characteristics of the pupils, the training requirements and the manner in which the results are to be processed and reported.

For each of the examples that follow, the source, ie. the test collection or agency from which it has been selected, is given and discussed; this is followed by a specification of the example, followed by the example itself. For some of the integrated assessments, the complete task is described, for others only a summary is given.

Material from some of the newer initiatives in assessment is also included: the Oxford Certificate of Educational Achievement (OCEA); and the various London graded assessments projects: GAIM, (Graded Assessments in Mathematics); GASP (Graded Assessments in Science Project); and GAML (Graded Assessments in Modern Languages, London).

The OCEA material is given as an example of the records of achievement initiative, being one of the nine pilot schemes supported by the Department of Education and Science. The aim is to assess a wide range of pupils' achievements and experiences. Students participate actively in the learning process through the establishment of specific negotiated short-term goals, which tends to improve motivations. Only a flavour of the many possible assessments can be given, in three subject areas (Examples 17 to 19).

The graded assessment projects are complete assessment schemes in which the pupils must meet specific criteria related to different levels of attainment. Pupils are assessed only when they are ready to tackle the range of tasks at a

particular level. There is an emphasis on setting tasks which are likely to be relevant to the pupil's own experience and to have practical applications and value to adult and working life. This leads to a great variety of assessment tasks of which only one is included here (Example 20).

The final illustration, Example 21, is given as a possible task which could provide assessments of science, language and mathematics on the basis of the same piece of work undertaken by an 11 year old pupil.

Table 1

Tests composed of
discrete questions

Can sample from a wide
variety of knowledge, skills
and understanding.

Can make it difficult to assess
the integration of knowledge
and skills.

Can be artificial.

Can be easy to administer.

Can be administered in a
relatively short period of time.

Marking can be relatively objective
given suitable marking schemes
and training.

Marking can be accomplished
relatively easily.

Readily yields numerical data which
can be difficult to interpret
meaningfully.

Tests consisting of
integrated tasks

A single task may be limited to a
few aspects of knowledge, skills,
and understanding but may offer
detailed assessment of these.

Can allow the integration of
knowledge and skills.

Can be realistic.

Administration in a standard form
may need some training.

Can be time consuming.

Marking will be relatively
subjective unless procedures are
carefully set out and some training
given.

Marking can be difficult and may
need special training.

Qualitative data can be
diagnostically illuminative and are
easily gathered. Analysis of
responses can yield categories of
response and numbers of pupils in
each category.

Example 1

Source

Early Mathematics Diagnostic Kit
(Lumb and Lumb, 1986, NFER-Nelson)

This is a collection of test items designed to identify early difficulties in mathematics learning. Items consist of pictures and a question read aloud by the teacher. Some items also allow the use of apparatus such as Unifix cubes. The items are related to specific behavioural objectives and are linked to suitable remedial activities.

Specification

Subject area: mathematics

Age group: 4-8 years

Individual administration

Presentation mode: oral (teacher) and pictorial

Task mode: mental

Response mode: oral and written

Individual questions not aggregated.

Example



How many? (read out by teacher)

Note the highest number of objects the child can identify correctly without verbally counting.

Give child paper and pencil.

Write down the number. (read out by teacher)

Note how the child writes the number. Is the number formed fluently and quickly?

Example 2

Source

Maths 7

(NFER, 1987, NFER-Nelson)

This is a test designed to assess the mathematics attainment of children near the end of the school year in which they reach their seventh birthday. It is principally a broad monitoring test.

Specification

Subject area: mathematics

Age group: 6-7 years

Group administration

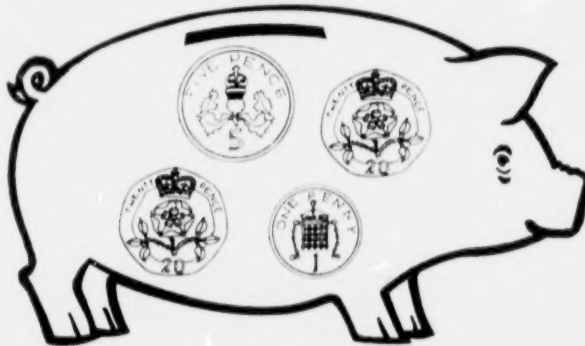
Presentation mode: oral (teacher) and pictorial

Task mode: mental and/or written

Response mode: written

Individual questions aggregated into total score.

Example



Answer _____ p

(Teacher reads) how much money is in the piggy-bank? Write down the answer on the line.

Example 3

Source

Maths 11

(NFER, 1984, NFER-Nelson)

This is a test designed to assess the mathematics attainment of children towards the end of the academic year in which they reach their 11th birthday. It is principally a broad monitoring test which is norm-referenced. Procedures are given for teachers to compare this performance of their own groups with the national facility value for each question.

Specification

Subject area: mathematics

Age group: 10-11 years

Group administration

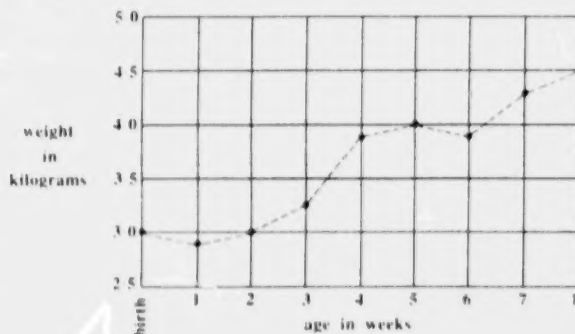
Presentation mode: written

Task mode: mental and/or written

Response mode: written

Example

The graph shows the weight of a baby girl in the first eight weeks of her life.



48. What was the baby's weight at birth?

Answer _____ kilograms

49. How heavy was the baby when she was 8 weeks old?

Answer _____ kilograms

50. How old was the baby at the end of the week in which she gained most weight?

Answer _____ weeks old

Example 4.

Source

Special Access Mathematics 1A
(NFER, 1986, NFER-Nelson)

This is a multiple-choice test restricted in use to LEAs who use it for monitoring exercises. It is norm-referenced and consists of discrete questions designed to test the understanding of mathematical concepts and processes or to test the use of mathematics in everyday situations. The distractors (wrong options) were selected to reflect procedural errors commonly followed by children of this age and misconceptions often held by them; that is, they provide diagnostic indications.

Specification

Subject area: mathematics

Age group: 10-11 years

Group administration

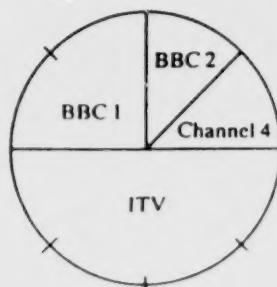
Presentation mode: written

Task mode: mental and/or written

Response mode: multiple-choice

Example

This pie chart shows the favourite TV channel of 1600 people:



How many people like Channel 4 most?

A. 45

B. 200

C. 320

D. 400

E. 90

Example 5

Source

Chelsea Diagnostic Mathematics Tests

(Hart, Brown, Kerslake, Kuchemann, Ruddock, 1985, NFER-Nelson)

These are a series of tests, each concentrating on one topic within the secondary school mathematics curriculum. They were designed as diagnostic instruments to be used for ascertaining levels of understanding and for identifying the incidence of errors. The item shown is from the measurement test.

Specification

Subject area: mathematics

Age group: 11+ to 14+

Group or individual administration

Presentation mode: written

Task mode: mental and/or constructions (depending on child)

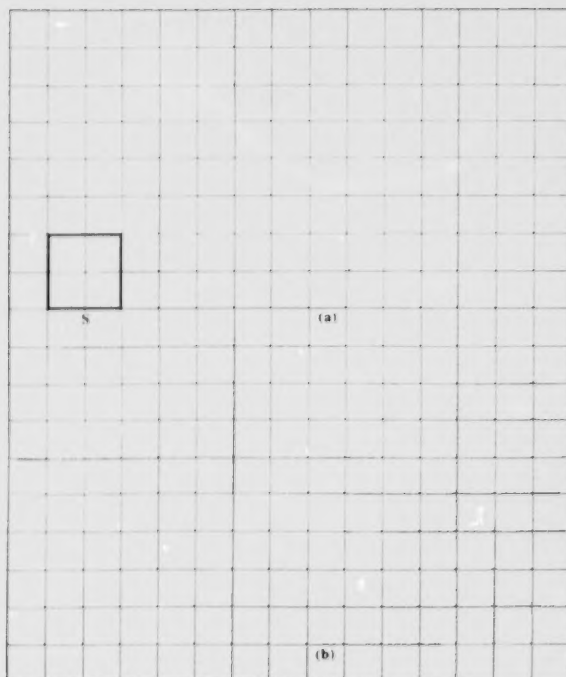
Response mode: construction

Example

17. On the squared paper below, draw

(a) A square whose **area** is **twice** (double) the area of square S.
Put a cross on (a) if you think it is impossible

(b) A square whose **perimeter** is **twice** (double) the perimeter of square S. Put a cross on (b) if you think it is impossible



Example 6

Source

Item from tests developed for Graduated Tests in Mathematics for the Lower Attainers in Secondary School Feasibility Study.

This project developed a series of tests for use with the bottom 40% of pupils in secondary schools. Assessment materials in a variety of modes were investigated, including written, oral, practical and investigative modes. One task investigated was a realistic simulation of shopping in order to test the Cockcroft Foundation List target "Recognise coins and notes, and know that 100p = £1. Handle money with confidence".

Specification

Subject area: mathematics

Age group: 13-15 years (low attainers)

Individual administration with assessor

Presentation mode: oral

Task mode: mental

Response mode: oral and performance

Example

The pupil is given £7 (a £5 note and two £1 coins) and is required to buy a series of articles, check the change and state how much money is left. After some purchases the wrong change was given by the assessor and the pupil's reactions to this noted.

Example 7

Source

Maths with the Micro Survey: APU 1987

This is a test developed for the APU 1987 secondary mathematics survey. The pupil has to plan a weekly budget using a Spreadsheet package. The pupil has to adjust the budget to fit given amounts of pay using the Spreadsheet. The item gives opportunities to examine understanding of percentages and of changes in percentages given a new base quantity. The microcomputer allows amendments and corrections to be made flexibly and the realistic situation provides a vehicle in which to examine the application of mathematics to everyday life.

Specification

Subject area: mathematics

Age group: 15-16 years

Individual administration via assessor and microcomputer

Presentation mode: oral and microcomputer

Task mode: mental and keyboard

Response mode: oral and microcomputer

Example (see source explanation above)

Screen layout during a typical run.

BUDGET		PAY	28.00	50.00 NEW PAY			
GUESS%	REAL%	COST	NEEDS	NEWCOST	GUESS%	REAL%	CHANGE%
20.00	17.86	5.00	FARES	5.00	SAME	10.00	-7.86
16.00	14.29	4.00	LUNCH	4.00	SAME	8.00	-6.29
40.00	35.71	10.00	HOME	15.00	UP	30.00	-5.71
28.00	25.00	7.00	FUN	20.00	UP	40.00	15.00
8.00	7.14	2.00	HOLS	6.00	UP	12.00	4.86
112.00	100.00	28.00	TOTAL	50.00	SAME	100.00	0.00

Example 8

Source

Diagnostic Test in Addition and Subtraction (NFER, 1987, unpublished)

This test uses the micro-computer as a diagnostic tool to build a picture of the misconceptions a child has when attempting certain arithmetic procedures.

Questions are generated by the program to test hypotheses and determine which particular misconception(s) the child is following. The results are reported as an error category rather than as a score.

Specification

Subject area: mathematics

Age group: 7-11 (children having difficulty with addition or subtraction)

Individual administration via microcomputer display

Presentation mode: microcomputer display

Task mode: written

Response mode: keyboard entry

Example

Two of the largest set of questions involved are follows in the form of the actual computer print out. The pupil response has been entered in, and the comment on the right is a diagnostic interpretation produced by the programme from the evidence of the pupil response.

Questions and answers:

Question 1: 970

- 518

468

Smaller from larger

0 - N = N

Question 2: 804

- 727

100

No decomposition/zero answer

After a test in which a larger number of such items are attempted by any one pupil, the programme gives a diagnostic summary for the teacher. An example is produced below. The asterisks indicate the most significant features.

FREQUENCY OF OCCURRENCE	ERROR
*** 7	Smaller from larger
* 2	Number bonds
** 7	No decomposition/zero answer
1	$0 - N = N$
2	$0 - N = 0$

Incorrect answers are probably due to the following error(s):

Smaller from larger

No decomposition/zero answer

Example 9

Source

Integrated task devised for 7 year olds by APU Mathematics team (untrialled)

Specification

Subject areas: mathematics and language

Age group: 7 year olds

Individual administration or small groups

Presentation mode: oral

Task mode: practical and mental

Response mode: oral and written

Example

1. Apparatus

200g bag of red kidney beans

Balance

Containers of different sizes

Measuring jar or long cylindrical container

Calculator

2. Task

- i. Estimate the number of beans in a 200g bag of red kidney beans.
- ii. Using any of the apparatus provided make a better estimate without counting out all the beans.
- iii. Check using a different method from ii.
- iv. Pupils to record results and to give an oral account of what they did to obtain their results.

3. Possible methods of tackling the task

- i. Overall strategy
Count the number of beans in a small unit; relate the smaller unit to the whole.
- ii. Possible tactics
 - a. Use balance successively to halve number of beans until a manageable number of beans is obtained; multiply up to obtain number in 200g.
 - b. Count the number of beans in a small container; find the number of small containers which will hold the 200g of beans.
 - c. Pour the 200g of beans into the long cylindrical container. Measure the height; now pour out most of the beans and measure the height of a proportion of them; relate the smaller height to that of the 200g.

4. Assessment

- i. Overall strategy
Could be assessed on a rating scale. Evidence for pupils' appreciation of the strategy would be obtained from their recording and oral account of what they did. Levels of mathematical argument used could be similarly assessed.
- ii. Carrying out the task: use of mathematical skills:
 - a. Accuracy of counting
 - b. Accuracy of calculation
 - c. Correct choice of operation
 - d. Understanding and use of balance
 - e. Calculator skills: accuracy of keying, order of operations, interpreting display
- iii. Communication skills
 - a. Recording
 - b. Oral

Example 10

Source

NFER Reading Ability Series Level A
(NFER, in press, NFER-Nelson)

This is part of a new series of reading tests which assess reading as a purposeful activity using realistic material which might be encountered within or outside the classroom. Different types of text both narrative and expository are included and the children's response to these assessed through cloze passages and open-ended questions.

Specification

Subject area: language

Age group: 7-8 year olds

Group administration

Presentation mode: written

Task mode: written

Response mode: written

Example

Exercise 1

Please look at this notice, and then answer the five questions

LOST:
2 kittens
one ginger, one black & white

If found, please contact
Mrs. J. Blake,
33 Hillside Lane,
Pipsam, Lanes
Tel: Pips. 77369

£10 reward
for
each kitten

What colour are they? _____

What would you get for finding both kittens? _____

Suppose that you found the kittens, in what two ways could you let Mrs Blake know? _____

1 How many kittens were lost? _____

2 Who lost the kittens? _____



Example 11

Source

Assessment of Speaking and Listening (Oracy)
taken from APU surveys, 1982-83.

These listening and speaking tasks were administered by an assessor working with pairs or groups of pupils seated round a table. In this particular task, two pupils listen to a taped description of how a spider builds a web; then, using a sequence of pictures, they are required to tell another pair of pupils about the process.

Specification

Subject area: language

Age group: 11 year olds

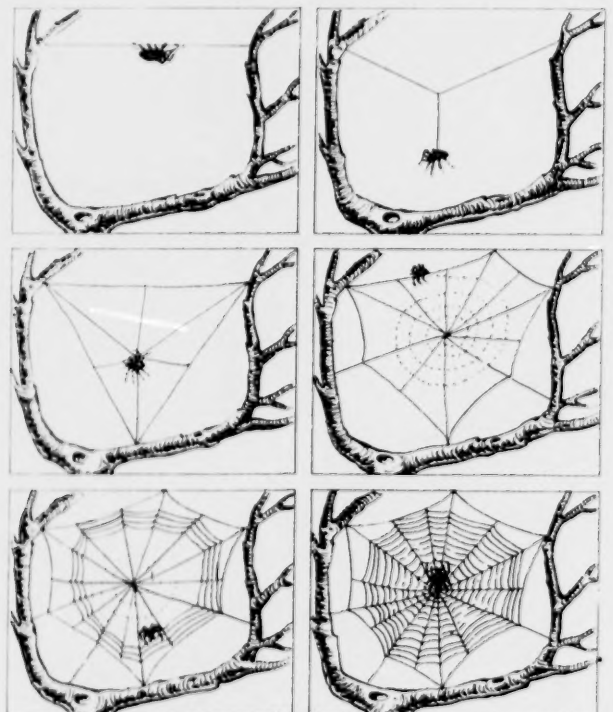
Individual or small group administration

Presentation mode: aural (tape) and pictures

Task mode: oral

Response mode: oral

Example



Example 12

Source

Assessment of Speaking and Reading

Task taken from APU age 11 oracy survey (1983)

Specification

Subject area: language

Age group: 11 year olds

Administration: small group

Presentation mode: reading and aural

Task mode: oral

Response mode: oral and written

Example

This is a task for two children and involves three basic activities for assessment.

These are:

1. Extracting information from a simple text.
2. Exchanging that information.
3. Reaching consensus through collaborative discussion.

Each child is given information about one of two fictitious plants (the Kava and the Bulu) presented in comic strip style. The children are asked to read the information and note down the merits and demerits of their allotted plant with respect to its suitability for planting on an island which has a range of given agricultural problems.

Having completed this task, they then proceed to tell their partner what they have found out about their plant. This exchange leads on to discussion about

the advantages and disadvantages of each and a final decision about which should be cultivated on the island.'



The soil has to be dug well, to grow the kava plant.



A lot of compost has to be added to the soil to feed the plants.



Once the plants have started to grow, they need very little water.



The plants grow stronger if the weeds are not allowed to flourish.



The Kavas are resistant to disease and pests.

the kava



When ripe the kavas are difficult to harvest, as the leaves are very prickly.



They need a lot of preparation to cook them.



They are extremely tasty vegetables.



The seeds are very cheap.



The seeds have to be planted every year.

Example 13

Source

Assessment of Writing

Tasks taken from APU surveys (1979-83)

The survey framework included extended pieces of writing with relevance to different curricular areas, as well as shorter writing tasks such as editing text produced by pupils of the same age, commenting on reading passages, and making notes for specific purposes. Other tasks required the integration of writing with diagrams or illustrations, as in the preparation of posters.

The writing is assessed using both impression and analytic scales, applied by teachers of the relevant age groups.

Specification

Subject areas: language and writing

Age group: 11 year olds

Group administration

Presentation mode: written and oral

Task mode: written

Response mode: written

Example

	THE BALLOON GAME	
	A group of people who all think they are very important to the rest of us	
	are riding in a balloon. They are a farmer, a cook, a teacher, a social	
	worker, a prime minister, a nurse, an actress and a miner.	
	The trouble is that some of them have to jump out or the balloon will not	
	fly high enough. Whoever can argue the most convincingly will stay longest	
	in the balloon. Decide who <u>you</u> think is really the most important person	
	of the eight and write a speech for him/her which will convince the others	
	that he/she can stay in the balloon.	

The purpose of this task is to assess pupils' ability to write persuasively and from a particular point of view.'

THE EARLIEST THING I CAN REMEMBER

| Think of the earliest thing you can remember when you were very young. It |
| could be a place, a person, something that happened, even an object. |
| Describe it as clearly as you can. |

The purpose of this task is to assess pupils' ability to produce a short piece of descriptive writing.

Example 14

Source

Assessment of Speaking and Listening (Oracy)

Task taken from APU age 11 oracy survey (1983)

Specification

Subject areas: language and science

Age group: 11 year olds

Administration: part individual, part small group

Presentation mode: aural and practical (instructions from assessor to be followed by pupil)

Task mode: oral

Response mode: oral and written

Example

This task comprises a sequence of activities based on a scientific experiment similar in principle to "Newton's cradle". When a marble runs down a plastic track and hits one end of a stationary group of four, the marble at the other end of the group is impelled up the track, rises, slows and stops, and then runs back to repeat the whole process (see diagram):

1. the assessor instructs pupil A on how to carry out the experiment and to observe what happens;
2. then pupil A similarly instructs pupil B;
3. pupil B describes his/her observations to the assessor;
4. both pupils discuss the experiment and develop a hypothesis to explain their observations;

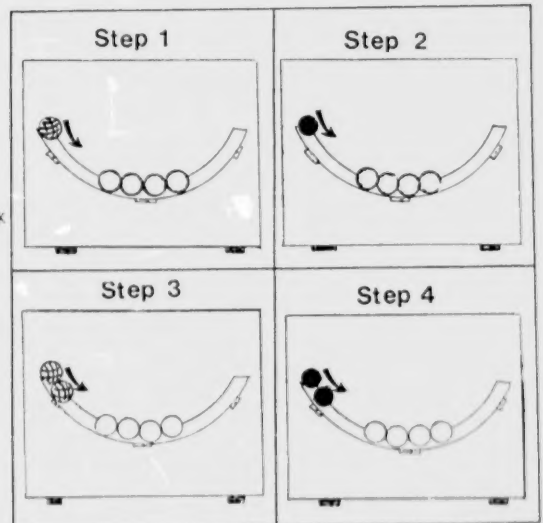
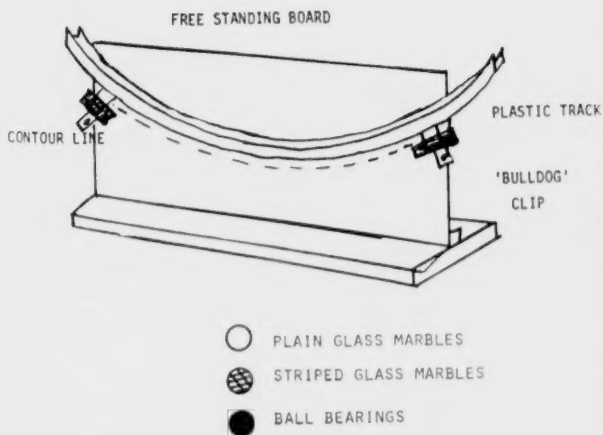
5. pupil A states this hypothesis to the assessor;

6. two more pupils are brought in, and produce a written account of the experiment, with diagrams.

Assessment

Pupil A: instructing Pupil B (2), hypothesising (4), reporting the hypothesis (5) and the written outcome (7).

Pupil B: describing observations (3), hypothesising (4), informing (6) and written outcome (7).



Example 15

Source

Task taken from APU German Test

The APU Modern Languages framework for assessment involves tests of listening, speaking, reading and writing. All are set in realistic contexts and the pupils are asked to carry out tasks which would be appropriate within those situations. An example of an item in a listening test is illustrated below.

Specification

Subject area: foreign language

Age group: 13 year old learners of German as a first foreign language.

Presentation mode: oral

Task mode: oral

Response mode: selection of correct picture.

Example

You are talking to another camper. She tells you what her camping holiday has been like.

You are talking to another camper. She tells you what her camping holiday has been like

An einem Tag war besonders mieses Wetter. Nachts lagen wir in unseren Schlafhäcken im Zelt. Da hat's begonnen in Strömen zu regnen. Wir mußten unsere Regenmäntel und Gummistiefel anziehen, um unser kleines Zelt zu sichern. Ich mußte die Lampe halten und mein Mann hat die Heringe tiefer eingeschlagen. Wir waren bis auf die Haut nass und hatten unheimlich schlechte Laune.



Example 16

Task from Piagetian Reasoning Tasks (previously CSMS Science Reasoning Tasks)
(Wylam H and Shayer M 1978)

These tests are designed to assess the ability of children and adults to use concrete and formal reasoning strategies. The tasks were intended to be used as an aid to curriculum development in science teaching. In the task shown below the pupil is asked to say how he would plan tests to distinguish the effects of three variables: how the length, weight and push of a pendulum determine the period of oscillation. Only the length is important, but the student must design an experiment controlling the appropriate variables in order to reach this conclusion.

Specification

Subject area: science

Age group: 9 to 16 years

Individual administration

Presentation mode: written and diagrammatic

Task mode: practical

Response mode: written

Example (see source above)

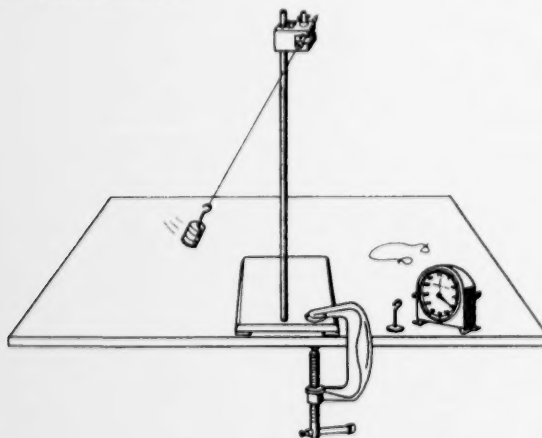
Equipment

Stopclock or watch with second hand

2 weights (say 100 and 400 gram slotted weights on hangers)

2 strings looped at either end. (Loop to loop distances of 69 cm and 35 cm will give about 17 and 22 swings in half a minute. The numbers are not critical, but try to avoid lengths giving either 15 or 20 swings.)

Firm support to hang pendulum from.



*For information on the use, development, statistics etc. of this Task see Science Reasoning Tasks Booklet, NFER, Windsor.

Example 17

Source

OCEA Mathematics

The OCEA teacher's guide to mathematics, "Putting it into Practice", gives 57 sets of assessment materials, together with illustrative solutions. The materials are divided into short problems, investigations, practical problems, quick response questions and projects. The example shown is one of the practical problems.

Specification

Subject area: mathematics

Age group: 13-15

Individual task

Presentation mode: written with illustration

Task mode: practical

Response mode: construction

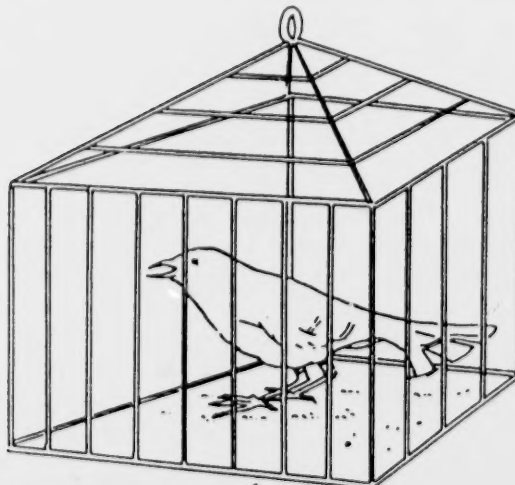
Example

BIRD CAGE COVER

Sketch a free-hand plan for a cover for this bird cage.

Choose a scale to fit your paper and draw a scale plan for the cover.

Use your plan to make a model of the cover in paper, to see if it works.



Example 18

Source

OCEA Modern Languages

The OCEA Modern Languages assessment framework is built on the four skills of listening, speaking, reading and writing. Detailed prescriptions are given for the assessment of these skills and combinations of them. These are largely realistic simulations of relevant situations. One of these, to assess talking, is shown below.

Specification

Subject area: modern languages - French

Age group: 13-15

Small group task.

Presentation mode: written

Task mode: oral

Response mode: oral

Example

Topic: Shopping for a present

Context

Student A/teacher: You are a shop assistant in a French gift shop.

Student B: You are in a gift shop in France buying a present for a friend or relative.

Aim of task

Through exchange of information *Student B* buys a gift which he/she can afford for a friend or relative.

POUPÉES Costume breton 29F 00 Costume normand 30F 00	POSTERS Grands 15F 00 Petits 8F 00	T-SHIRTS Avec Image 23F 00 Avec motif 25F 50	A. You are a shop assistant in a French gift-shop. [STUDENT/TEACHER ROLE]
---	---	---	---

 →  11 ans	8 28 francs	You are in a gift shop in France buying a present for a friend or relative.
Elle collectionne les poupées		

Example 19

Source

OCEA Science

The OCEA Science assessment framework is based on four processes; planning, performing, interpreting and communicating. Each of these is linked to a small set of skills; for example, planning is linked to (1) producing testable ideas and (2) designing investigations.

These are then further divided into criteria which are positive statements describing achievements within the skills. In all there are 34 criteria and an activity can be used to assess a variety of these. The student activity shown below could be used to assess one skill associated with producing testable ideas, that the pupil can suggest other ideas related to those already investigated, but it could also be used for many other skills.

Specification

Subject area: science

Age group: 11-14 years

Individual task

Example

Presentation mode: written

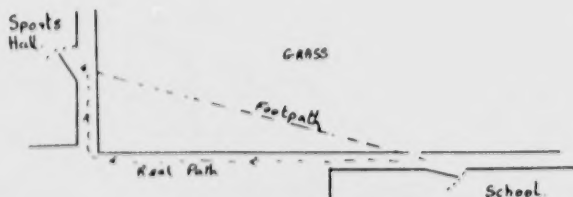
Task mode: practical and written

Response mode: practical and written

Worksheet: Footpath

Many students take a short cut across the grass to get to the sports hall.

Map



Balbir said that the grass would be thinner where people had been walking. She thought of a way of finding out if it was true.

- 1 Think of something else that might be different about the path and write it on your sheet.
- 2 Think of an experiment to see if you are right and describe it on your sheet.
- 3 Do your experiment and then write about

Example 20

Source

The Graduated Assessments in Mathematics scheme ILEA, WLSEB and King's College London is similar to OCEA's in that it provides a categorisation of topic criteria which can be assessed in many different ways. The pupils' progress on particular criteria can be assessed through classwork, self-assessment or through review sheets which may be undertaken as a final check after all the topic criteria in one area have been recorded as satisfied. Some assessments arise from group activities concerned with a practical problem. One of these is shown below.

Specification

Subject area: mathematics

Age group: age 11 upwards

2-3 students in group or independently

Presentation mode: written

Task mode: oral and written

Response mode: written, diagramatic

Example

P199

Breakfast Cereals



1. What is the most popular breakfast cereal in your class?
2. Which cereals give best value for money?

Your group must decide how to find the answers to these two questions.

- Do you need more information?
- How will you find it out?
- What jobs need to be done?
- Who will do what?

Discuss and decide how you will present your results.

Example 21

Source

SAILBOATS

Integrated tasks devised for eleven-year-olds by APU teams (untrialled in this version - used in an earlier version in science monitoring work). This task is rather similar to those in Appendix E and could be used as part of the collection in which they might (after full development) be made available.

Specification

Subject area: science, mathematics and language

Age group: 11 year-olds

Small group administration

Presentation mode: oral and written

Task mode: practical, oral, written and mental

Response mode: oral and written

Example

Aim of the Experiment

To discover the most effective shape of boat given

1. a choice of two boats;
2. two different types of sail;
3. a third sail of given shape but choice of size to be decided by the pupil.

Apparatus: Tray 60cm x 40cm x 3cm

2 different shapes of "boats"

2 split pin type masts

Minifan with stand

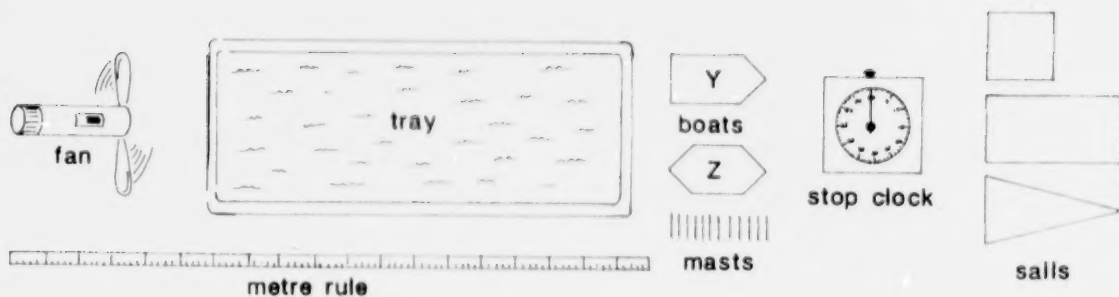
Metre rule

Stop clock (marked in one-tenth second divisions)

Small bucket

Plastic sheet

The equipment should be laid out on the plastic sheet as illustrated below:



Administration

The task involves collaboration between three pupils. The pupils are told that the problem involves finding the best shape of sail for some model boats. The assessor shows the pupils how to fit sails to the model boats and how to use the other equipment. The pupils are told that they are to record how long the boats take to sail the length of the tray or part of it. The assessor then ascertains that the pupils understand the task.

One of the three pupils is given charge of the experiment. The pupil in charge is also in control of the fan and release of boats. The other two have, respectively, the tasks of timing the boats with the stop-clock and recording the experiment.

Two sets of sails are provided of different sizes.

1. Each of the pupils is asked to make drawings of the sails and of the boats and to record the dimensions on the drawings (including the height).
2. They are told to find out the area of the sails and boats.
3. The three pupils are then asked to discuss how they will carry out the investigation, and
4. to record on their work sheet the steps they have agreed to take in the experiments.
5. The pupils are then asked to discuss which boat and sail they think will give the fastest speed. They are not at this stage prompted to identify what factors relating to the size and shape of the boats and sails are important in determining this.
6. If after ten minutes pupils have not referred to any of the measurements of the boat in their discussion they are prompted to consider taking the following ones into account:

- a. the dimensions/shape of sails
- b. the dimensions/shape of boats
- c. the height of the sail
- d. the area of the sail.

7. The pupils are then told to carry out the series of experiments and that each of them in turn will act as time-keeper and as recorder.

Reporting Back

8. The pupils are asked to complete a report sheet giving their results and to comment on how their findings compared with their predictions.

Classroom Follow-up

Based on their findings all pupils in the class could investigate sailing ships and discuss their suitability for various applications, speed and manoeuvrability. They could also carry out further experiments investigating the effects on speed of varying the dimensions of the sails.

Assessment

- * Ratings of effectiveness of group discussion and oral presentation (hypothesising, predicting, explaining) (3, 5).
- * Assessment of the clarity and appropriateness of features of the written records (sequencing, tabulation, recursivity) (4, 8), plus:
 - checklist of carrying out of mathematical operations (1, 2)
 - checklist of carrying out scientific operations (5, 6, 7)

Acknowledgements

Acknowledgements and thanks are due to the following organisations for permission to use their material.

National Foundation for Educational Research

NFER-Nelson Publishing Company

Oxford University Delegacy of Local Examinations for OCEA material

University of London School Examinations Board for Graded Assessment material

Science Reasoning, Kings College, University of London

Assessment of Performance Unit

NATIONAL TESTING AND EQUAL OPPORTUNITIES

Submission by the Equal Opportunities

Commission to the DES Task Group on

National Testing

December 1987

Table of Contents

1.	Summary	1
2.	Introduction	3
3.	Evidence of Gender Differences	3
3.1	Language Differences	3
3.1.1	Reading	3
3.1.2	Writing	4
3.1.3	Test Format	4
3.1.4	Motivation and Context	4
3.2	Mathematics Differences	4
3.3	Verbal and Non-Verbal Ability Differences	5
3.4	Science Differences	5
3.5	Summary	6
4.	National 'Benchmark' Testing	6
4.1	The Use of Test Results	7
4.2	Adjusting for Gender Differences	8
4.3	Gender-Fair Tests	9
5.	Further Work	9
6.	References	11

1. Summary

1. There are established differences in test performance between girls and boys throughout the age range 7 to 16 and across different areas of the curriculum. Differences in performance at various ages have been shown in language, mathematics, science and verbal and non-verbal reasoning (section 3).
2. Various aspects of test design have differential effects on girls' and boys' performance. The use of practical tests, the form of questions adopted (eg. multiple choice vs. essay questions) and the context and content of individual test items have all been shown to affect the relative performance of girls and boys (3.1.3, 3.1.4, 3.2 and 3.4).
3. These findings have important implications for any proposed national system of testing pupils at 7, 11 14 and 16:
 - (i) Where the results of a number of tests are combined to give a single overall test score the resulting gender difference will reflect the weighting given to the component test areas (3.5).
 - (ii) If test results are to be used to compare schools, it is important that the proportion of pupils of each sex in the relevant age group within each school should be known and taken into account. The effects of gender differences on inter-school comparison will, of course, be greatest where the comparison involves a single-sex school (4.1).
 - (iii) If test results are used to allocate individual pupils to ability groups for particular subjects, differences in test performance will affect the gender composition of different groups. If scores from a number of tests are combined and used for general banding or streaming purposes, the gender balance of each band or stream will be dependent on the weighting given to the different component tests (see 3(i) above) (4.1).

- (iv) Tests can be constructed to reflect any desired degree of gender difference, or none. There are no generally accepted objective external criteria for test construction which can guide a test constructor in choosing the relative weightings for different test questions or topic areas. If tests are constructed, on grounds of equity, to give equal average scores for girls and boys, consistency would require that this should be done at all ages. This would raise questions about other assessment procedures such as GCSE (4.1 and 4.3).
- (v) An alternative to eliminating gender differences in average test performance by revising the construction of the tests, is to standardise test scores according to different gender norms. In the EOC's view this would constitute unlawful discrimination (4.2).

2. Introduction

This report is divided into two parts. The first gives a brief summary and evaluation of the evidence for gender differences and the second part looks at some possible consequences of these in a system of national 'benchmark' tests. The areas studied are those of language, mathematics, ability tests and science.

Because of time limitations, only a very general coverage of the issues is possible. Nevertheless, because the subject is complex, touching upon basic issues of educational achievement, a more extensive treatment is desirable. In a final section an outline is given for such further work.

3. Evidence for Gender Differences

3.1 Language Differences

3.1.1 Reading

In terms of vocabulary, boys show greater knowledge in Primary school (Douglas et al., 1968) with a persisting difference through the secondary school years (Wittig and Peterson, 1979).

In reading comprehension, an early (7-8 years) advantage in favour of girls, becomes very small by the age of 11 years and by the age of 16 years there is a small advantage to boys (Fogelman et al., 1978; Davie et al., 1972; Douglas et al., 1968; NCDS, 1972).

The American National Assessment of Educational Progress (NAEP, 1986a) reports superior overall performance for girls in reading up to 17 years and the British Assessment of Performance Unit (APU, 1982) does so at ages 11 and 15 years.

3.1.2 Writing

According to NAEP (1986b) girls do better than boys overall at all ages on writing tasks, and this finding is supported by the APU results at 11 and 15 years. NAEP divides writing into 'informative', 'persuasive' and 'imaginative.' In all types of writing girls perform better than boys at all ages from 9 to 17 years.

3.1.3 Test Format

There is considerable evidence that, relatively, boys do better than girls in multiple choice questions when compared to free response or essay questions. Murphy (1980), dealing with geography O' level exams, finds an increased pass rate for boys when the percentage of multiple choice questions is increased. Wood (1978) reinforces this for language examinations.

3.1.4 Motivation and Context

Wood (1978) provides evidence, based on O' level English language exams, that boys do better than girls when the topic of a question is a 'masculine' one (e.g. trains) and girls do better when it is a feminine topic (e.g. a story about a young girl's feelings).

3.2 Mathematics Differences

At the end of primary schooling, boys and girls appear to have similar performances on, largely arithmetical, mathematics tests. By 15 years the boys show an advantage (Douglas et al., 1968). Wood (1976) shows that boys do better than girls in 15 years old public examinations in questions concerning measurement and spatial topics. The APU data (APU, 1986) confirms this at age 15 and also shows that there are only very small differences in number skills and modern algebra. Boys also tend to do better in applied and practical mathematics. At the

age of 11 years boys are more confident in measurement and practical tasks than girls, and at this age the girls performed better than the boys only on computation tasks. There was only a very small difference at this age again in modern algebra. There is also a suggestion that there is only a very small difference at 11 years in problem solving tasks.

Plake et al. (1982), using university students, found that the arrangement of items in a mathematics test was important. With a traditional easy-hard ordering males did better than females but for an arrangement where easy and hard items were uniformly spread across the test these differences disappeared. There appears to be little data concerning such an effect in school age pupils.

3.3 Verbal and Non-Verbal Ability Differences

At the end of primary school, girls appear to score higher than boys on non-verbal tests and considerably higher than boys on verbal tests. By the end of secondary school, boys score somewhat higher on the non-verbal tests. (NCDS 1972; Douglas et al., 1968). Within the non-verbal domain, boys appear to begin to outperform the girls from about age 13 (Macoby and Jacklin, 1974).

3.4 Science Differences

The APU science monitoring programme (APU, 1982) shows that at the age of 11 years boys tend to do better than girls in the application of taught science concepts and in practical investigations. At age 15, results have to be interpreted with care since boys generally have had more exposure to science. The boys do better than girls at this age in the use of equipment, interpreting data, reading information and applying physics concepts. The latter difference is also apparent at age 13.

3.5 Summary

An important feature of gender differences is that context, content and format of test questions can affect these differences, even to the extent of reversing a difference on otherwise similar questions. The multiple choice format appears to favour boys as does practical testing.

4. National 'Benchmark' Testing

It is proposed by the Government that nationally prescribed tests will be administered and marked by teachers at the ages of 7, 11, 14 and 16 years in the areas of mathematics, language and science. It is possible that these ages could be varied somewhat, but that would not affect the substance of what follows. At the time of writing there are no details of how the test scores will be used, but some general indications are available.

The national curriculum consultation document refers to the use of the tests for providing parents with information for comparing schools, and since it is intended that parents will be given details of their children's test results, it is likely that these results will also affect within-school decisions on ability grouping etc. There have been suggestions that the test results will be useful as 'diagnostic' instruments for the teacher and pupils. It is unlikely, however, that the tests will be detailed or sensitive enough for this purpose and most professionals would not accept that a single instrument properly can combine a 'monitoring' and 'diagnostic' function.

The following sections discuss the relevance of gender differences to tests which are designed for purposes of comparison rather than diagnosis.

4.1 The Use of Test Results

It is difficult to predict how the dissemination of test results to parents will affect their actions and perceptions. In terms of local

or central government actions, however, there are several clear possibilities. A LEA could, for example, use test scores as indicators of need and direct resources to schools with poorer performances. Alternatively, it might take the view that poor performing schools should become, say, candidates for closure or amalgamation.

Gender differences will be an important factor where there are disparate proportions of girls across schools, and of course in the case of single sex schools. If the tests produce, for example, lower average scores for girls, then those schools with high proportions of girls will tend to produce lower mean test scores. If such schools 'suffer', either through parent or LEA action then it could be said that gender discrimination was occurring, since relatively more girls than boys would 'suffer' as a result. One way to avoid such an outcome would be to carry out a valid statistical adjustment for gender differences when comparing schools. This will be elaborated upon in the next section.

The other relevant aspect of the use of test results is in individual selection and placement. Thus, a secondary school might have ability groupings for its classes, say for mathematics from the second year. If national mathematics tests scores are available on each child, there would be some pressure, presumably, for these to be used in the ability grouping procedures. In some cases schools may find it difficult to carry out ability groupings in any other way when pupils and parents have access to each child's test results. As pointed out in section 2.5, tests can be constructed generally to reflect any desired degree of gender difference (or none at all) with consequent effects on the numbers of girls and boys in different groupings.

It is generally accepted that there are advantages which follow from belonging to a high ability grouping, so the possibility of discrimination again arises. The problems of constructing gender-fair tests is discussed below. The principal issue here is that there are no generally accepted objective external criteria for test

construction which can guide a test constructor in choosing the relative weightings for different test questions or topic areas. If it were argued, on grounds of equity, that tests should therefore be constructed to give equal average scores for boys and girls, then consistency would require that this should be done at all ages. Furthermore, such a procedure would raise questions about other assessment procedures such as the GCSE and whether attempts should be made to balance results by gender more generally in relevant public exams.

It is worth pointing out that this issue cannot be resolved by appealing to historical precedents in test construction. What has occurred previously partly reflects the culturally related expectations of test constructors concerning gender differences (Gould, 1981).

4.2 Adjusting for Gender Differences

If gender differences are eliminated, in the manner suggested above, or by standardising scores according to separate gender norms, then differences between schools would not be expected to depend on the proportions of girls in the schools. If, however, gender differences in test scores are allowed to persist, then a complicated statistical adjustment would be necessary. A discussion of the problems surrounding such procedures is given by Goldstein (1987).

Suffice it to say here that the use of individual children's test results is required rather than school or LEA averages. It should also be noted that there are other factors, such as ethnic background social class and attainment at time of entry to school, which need to be adjusted for in order to make fair comparisons between schools or LEA's.

It seems that the complexities of the process effectively would rule out its routine use. Certainly it would require a level of expertise unavailable to most LEA's. Moreover, even if such adjustments could

be carried out routinely, similar arguments concerning consistency apply here as in the case of test scores used to compare and group pupils.

4.3 Gender-Fair Tests

It is now commonly accepted good practice that test constructors have an obligation to ensure that their instruments contain no sexual or racial stereotypes or any material which could be offensive to a particular group. There exist guidelines (American Psychological Association et al., 1985), and any national system of testing should aim to follow these broadly. It would also be important that the details of test construction are publicly available so that public scrutiny and challenge is possible if guidelines are not followed.

Of course, even where such guidelines are followed, these cannot be expected to eliminate all gender difference, so that their adoption does not remove the necessity for considering the issues of sections 4.1 and 4.2.

5. Further Work

The review of the evidence of section 3 has not attempted to evaluate existing research findings in terms of their quality. Some studies have been excluded on the grounds of unrepresentativeness, and it is clear that a more thorough review and evaluation of findings, especially concerning narrowly defined domains of achievement, would be useful. Such a review should also pay attention to quantifying differences, so that their relative importance can be assessed.

Some other issues are relevant also. The performance of boys and girls in single and mixed sex schools has been the subject of research, but the existing evidence does not seem to be clear-cut and needs careful evaluation. There have been several studies of teachers' expectations in relation to gender differences and this may well be important if teachers are allowed discretion in the administration and marking of tests. A review of public examination entry policies and results would provide further useful evidence on this issue.

Finally, whatever decisions are made about the implementation of national tests, it is extremely important that there is a proper, and independent, evaluation of their effects with respect to gender as well as other factors. Given the uncertainties about the likely effects, such an evaluation could at least be used to alert people to serious problems and possible violations of equal opportunity legislation.

6. References

1. American Psychological Association, American Educational Research Association, National Council on Measurement in Education (1985) Standards for Educational and Psychological Tests. Washington, D.C. A.P.A.
2. Assessment of Performance Unit (1982) Science in Schools. Survey Reports No. 1 at ages 11, 13, 15. London, DES.
3. Assessment of Performance Unit (1982) Language Performance in Schools. Survey Reports No.2 at ages 11, 15. London, DES.
4. Assessment of Performance Unit (1986) A Review of Monitoring in Mathematics, 1978-1982. London, DES.
5. Davie, R., Butler, N.R. and Goldstein, H. (1982) From Birth to Seven. London, Longman.
6. Douglas, J.W.B., Ross, J.M. and Simpson, H. (1968) All Our Future. London, Peter Davies.
7. Fogelman, K., Goldstein, H., Essen, J., and Ghodsian, M. (1978) Patterns of Attainment. Educational Studies, 4, 121-130.
8. Gould, S.J. (1981) The Mismeasure of Man. New York, W.W. Norton.
9. Goldstein, H. (1987) Multilevel Models in educational and Social Research. London, Griffin; New York, Oxford University Press.

10. Macoby, E.E. and Jacklin, C.N. (1974) Psychology of Sex Differences. Stanford, Stanford University Press.
11. Murphy, R. (1980) Sex differences in GCE Examination entry Statistics and Success Rates. Educational Studies, 6, 169-178.
12. National Assessment of Educational Progress (1986a) The Reading Report Card. Princeton, Educational Testing Service.
13. National Assessment of educational Progress (1986b) Writing. Princeton, Educational Testing Service.
14. National Child Development Study (1972) Second Follow Up of the NCDS. Report to the Social Science Research Council.
15. Plake, B.S., Anson, C.J., Parker, C.S. and Lowry, S.R. (1982) Effects of Item Arrangement, Knowledge of Arrangement, Test Anxiety and Sex on Test Performance. J. Educational Measurement, 19, 49-58.
16. Wittig, S. and Peterson, A. (1974) Sex Related Differences in Cognitive Functioning. New York, Academic Press.
17. Wood, R. (1976) Sex Differences in Mathematics Attainment at GCE Ordinary level. Educational Studies, 2, 141-160.
18. Wood, R. (1978) Sex Differences in Answers to English Language Comprehension Items. Educational Studies, 4, 157-165.

Acknowledgement

This submission is based on a research review undertaken by Professor Harvey Goldstein on behalf of the Equal Opportunities Commission.

TECHNICAL NOTE ON RELIABILITY, VALIDITY AND ERROR

Reliability refers to the scores assigned to pupils and is measured in terms of their stability over time or over different sets of equivalent tasks or over other variable assessment conditions. Computation of a reliability coefficient allows the estimate of an error of measurement: a statement of the range of fluctuation which is likely to be observed in a single pupil's score as measurement circumstances change.

Changes in the circumstances of measurement will usually have direct educational implications in national assessment. The formative purpose does not require that assessments are stable over long periods of time: quite the contrary, they should be sensitive to changes in the educational experience of the pupil. The definition of the national curriculum in terms of attainment targets will often mean that equivalence over different content areas is an objective of the national curriculum, not a prerequisite. For these reasons, arising directly from the differences between achievement and ability testing, we do not recommend standard errors of measurement as the most suitable basis for the computation of the confidence limits of a score.

Validity relates to an interpretation or use of test scores. It is obviously reduced by errors in the estimate of the scores themselves but is further reduced by unwanted influences on the relationship of the measures to the construct which they are interpreted to measure. We feel that the confidence limits of national assessment scores should reflect these influences which largely arise from the necessity of sampling achievements. Such additional sources of error are readily estimated and taken into account in generalisability coefficients (Cronbach et al, 1972; Brennan, 1983; Johnson, in press).

There are three other advantages of the generalisability approach. National assessment is to serve the needs of different users who will wish to interpret essentially the same data base in different ways. Different generalisability coefficients can be derived from the same data base according to the intended

interpretation. National assessments will produce pupil profile reports. Generalisability deals with the overall degree of confidence in a profile. Generalisability theory can be used to refine the data collection techniques in national assessment by showing how best to minimise threats to confidence with the resources available. It seems entirely appropriate that a formative system should use a formative approach to quality control, which is provided by generalisability theory.

Users will want to be able to make interpretations and draw conclusions with confidence but cannot be expected to be familiar with the translation of reliability or validity indices into confidence limits. We suggest that where appropriate these be reported directly, especially in the translation of standard assessment marks to levels and in evaluative reports on school performance. Of course, the analysis required to produce data for users will have to be undertaken by expert agencies, but they could then advise users on the degree of confidence that they could place in the results.

References

Brennan, R L. (1983) Elements of Generalisability Theory. The American College Testing Program.

Cronbach, L J, Gleser, G C, Nanda, H and Rajaratnam, N. (1972). The Dependability of Behaviourial Measurements. New York, Wiley.

Johnson, S (in press) Technical report on the APU Science Monitoring Programme. Department of Education and Science.

THE CONDUCT OF GROUP MODERATION MEETINGS

This account is speculative. It is intended to help readers to understand our intention in recommending group moderation. It is not intended to establish firm rules and procedures - which are for determination at a later stage. It is based on experience of the use of group moderation in public examinations.

1. Organisation

For the administration of national assessments, there will be need for a regional organisation between the SEAC and NCC and the schools which will act as a communication link. As moderation is the major channel of communication that link, referred to hereafter as 'region', will have moderation meetings as a prime focus of its attention. This report makes no recommendation about the body which should be responsible for this regional organisation.

2. A chairman should be appointed by the region to act as a link with moderation groups. It is possible that in the early years these could be full-time appointments with general responsibility for INSET especially at the primary level where there is little experience of formal assessment procedures. We have not designated a national assessment day. At age 7 certainly, and at age 11 probably, all assessment is likely to be distributed over time allowing one chairman to deal with a reasonably large number of groups and to carry standards and insights from group to group.

At other times of the year a full-time appointee can

- a. visit standardising meetings held within and for individual schools
- b. give direct training in schools in the use of standard assessment tasks.

Wherever there is more than a single teacher involved, schools should be required to hold an internal standardising meeting (see Section X) prior to the submission of internal assessments to the group moderation meeting. All

teachers responsible for the assessment of candidates are required to take part in an internal standardisation which has an important part to play in school self-evaluation and development. There will be a simple group meeting for age 7, and another for age 11. At age 14 there will be separate meetings for each subject. Similarly at age 16 there could be meetings to determine results for non GCSE pupils, but it is possible that the results for these could be considered together with those for age 14.

The meeting may be organised at the discretion of the school in order to suit its particular circumstances. Chairmen should notify centres of their intention to attend the internal standardising meeting.

The meeting should include activities which promote discussion about standards and enable the centre to achieve an agreed standard of marking. The purpose of a visit by a chairman is to give teachers opportunity to discuss not only the work presented but all aspects of the syllabus support material or administration that may require clarification. The exercise is designed to help communicate national requirements in terms of activities and standards.

3. The group will appoint a secretary to take care of incoming and outgoing documentation. The success of the system will depend upon the quality of documentation which will comprise:

- a. a special report form for each school which allows the display side by side of teachers' ratings, and of profile component scores derived from the standard assessment tasks. This should be preprinted every year so that raw scores on standard assessments can be lined up with level scores on teachers ratings.
- b. a pupil report form showing for each component the teacher's rating, the standard assessment level and a space for an agreed level report or an indication that follow-up assessment is required (Documents B).
- c. a form for reporting comments on the pedagogical, curriculum design and assessment implications of components discussed at the meeting.

Computer support would also be advisable. Transmission of data on disc will lead to rapid collation and dissemination of information. The investment in software will also facilitate the internal record keeping which is increasingly required for the initiation of Records of Achievement. Many schools already have arrangements with public examining groups to register their candidates by electronic means. Some generalisation of these arrangements is required.

4. Prior to the group moderation meeting

Groups of schools are established in such a way that each group forms a reasonably compact and stable geographical unit.

One document, Document A, will be prepared from the documents described in 3(a) above, aggregated by the region for the group and for all groups in the region.

The main aspect of interest is the extent to which levels in individual profile components have proved easier or more difficult to attain on the standard assessment tasks and tests. It can be anticipated that there will be a less than perfect match between teachers' ratings and the test results and it is that pattern of mis-match which is used in the moderation process to identify schools the assessments of which depart from expectation.

Decisions will be taken at regional level, on the basis of the region's aggregated results, on which profile components and/or standard assessment tasks require detailed attention at group meetings.

It is likely that at age 7 attention will focus on just the one of three assessment tasks which all in a group will have used in common, in order that the richness of contextual effects should not be swamped by an attempt to cover too many issues. At age 11 two assessment tasks should be covered in this way. There will probably be a three year cycle to ensure systematic curriculum coverage.

The group chairman will liaise with the group secretary to plan a format for the meeting that provides for the full discussion of the standard assessment material and of issues on the agenda likely to arise.

All members will receive personal copies of documents in the form of 3(a) above, for each school in the group, for the group as a whole and for the region as a whole.

5. Conduct of the meeting

Group moderation meetings are crucial in encouraging positive views on the syllabus and the assessment procedures. Thus the promotion of positive discussion of the materials, the drawing in of new and young teachers and the highlighting of principles and good practice are likely to facilitate not only the assessment procedures but also the continued success of the scheme. These meetings are the initial forum for the debate of issues and principles.

First phase

This will consist of a scrutiny by all present of document A from the group related to document A for the region.

The group will consider first at the regional level the relationship of teachers' ratings to standard assessment results component by component. Where there are sound reasons for a discrepancy between the two these will be recorded on a further document, document C, and a decision taken as to which is the better estimate of the component. Otherwise the standard assessments will stand.

The group will then compare both (a) the group's pattern of teachers' ratings and (b) the standard assessments with the agreed regional outcome. This consideration is of the two patterns of discrepancies across profile components, NOT of the absolute standards. Statistical moderation will highlight profile components requiring special consideration. Again any special reasons for departure from the regional pattern will be recorded. Otherwise it will be expected that the group follow the regional pattern.

Finally the group will consider the group's overall standards vis a vis those of the region. Again, statistical moderation taking into account demographic variables can serve to focus the group's attention and again any justification for departures from the regional standard will be considered.

By the end of this phase the group will be alerted to particular profile components where some adjustment by the individual schools within the group will be necessary to bring the group into line with regional standards.

In this phase discussion of the standard assessment material should provide teachers taking part with the opportunity to raise and discuss any other matters or problems relevant to the assessment. They may wish to seek guidance from the region on certain issues or to make proposals for amendments of amplifications to the conduct of national assessment. Matters raised in this way by groups would normally be discussed in the first instance by either the region or be referred, if necessary, to SEAC. The SEAC may also use the channel of the trial meetings to inform teachers of relevant matters or to request their views on aspects of the scheme of assessment.

Second phase

The group now considers each school's pattern of outcomes in relation to the expectations now established for the group as a whole. Sometimes all schools in a group may have to accept an adjustment to one or other component. That aspect will be emphasised and its assessment discussed in detail. At other times one or two schools may have been identified by statistical means as at wide variance in their marking and the chairman should, as tactfully as possible, invite them to comment on their assessments rather than their remaining in silent disagreement with the general feeling of the group.

It is essential that all teachers in the group have the opportunity to share their standards and experience of assessment with other teachers in the group. This process not only helps to establish comparability but also encourages staff development in assessment.

Under the guidance of the chairman the group may wish to discuss administrative issues connected with the submission of work and grades that might be helpful, particularly to new centres.

At the end of this phase the group must ratify adjustments school by school. These adjustments are in terms of percentages of pupils at a level. For example, an adjustment decision might require a school starting with 30% in

level 1, 50% in level 2 and 20% in level 3 to alter the distribution to 25% in 1, 50% in 2 and 25% in 3. It is for the individual school to determine which pupil assessments have to be changed to bring the school into line with the distribution approved by the group.

Third phase

The major patterns of individual pupil records will be considered and problems arising in relation to exceptional pupils, who will be indicated as such on Document B (see 3(b) above). The group will discuss the circumstances in which follow up assessments are required and notify the secretary of the need for follow up materials. Decisions on the first level assignments for individuals will however be the responsibility of the school.

All agendas will include standard items of business, such as the assessment or teaching or the administration or regulations of the national curriculum. These may arise from the agenda or from a paper such as the annual report of the chairman of chairmen or from items of concern to the group. The group chairman, in planning the format of the meeting, may consider inviting experienced members of the group to lead, or make structured contributions to, such discussions. A key aim in these parts of the meeting should be to promote the sharing of the best practice as well as full understanding of the national curriculum.

After the meeting

Each school submits amended Documents A and B to the chairman for ratification prior to reporting results to parents.

Group chairmen will be invited to an annual meeting at region to discuss issues which have been raised by the groups during the year. It is important that group secretaries record such matters in the minutes of the group so that they can be included in the agenda for the annual chairmen's meeting.

RECORDS OF ACHIEVEMENT

A record of achievement brings together the wide range of a pupil's achievements and records progress and activities both within and outside the classroom, including experiences and achievements reflecting qualities not tested by examinations.

The Government's consultation document: "the National Curriculum 5 - 16" recognises that the national curriculum will not remove the need to place a value on all of a pupil's achievements and to record and report these succinctly for the benefit of pupils, parents and employers. It states the Government's commitment to the national introduction of records of achievement in the following terms:

"The Government aims to set in place by 1990 national arrangements for the introduction of records of achievement for school leavers. Such records will have an important role in recording performance and profiling a pupil's achievements across and beyond the national curriculum (paragraph 32)".

Education Support Grants are being used to support development work in 9 pilot schemes involving 22 local education authorities and some 250 schools in England and Wales. The object of the schemes is to cast light on the best way to set about recording pupils' achievement with a view to establishing the greatest possible degree of agreement on crucial issues and to prepare the way for national guidelines.

Support for the pilot schemes has been provided over the three years 1985-88. A National Steering Committee was set up to oversee and monitor this development work. The schemes will make their final reports in March 1988 and the National Steering Committee will produce a final report in the Autumn of 1988. This will offer the Government draft national guidelines within which records of achievement can be introduced for all young people in secondary schools in England and Wales. This report will be published.

The National Steering Committee produced an interim report in November 1987, copies of which can be obtained from the Department of Education and Science, Publications Despatch Centre, Government Buildings, Honeypot Lane, Stanmore HA7 1AZ.

The evidence submitted by the National Steering Committee to the Task Group on Assessment and Testing is reproduced below.

RECORDS OF ACHIEVEMENT AND THE ASSESSMENT OF THE NATIONAL CURRICULUM:
EVIDENCE TO TGAT FROM THE NATIONAL STEERING COMMITTEE FOR RECORDS OF
ACHIEVEMENT

Introduction

1. Our aim in making this submission to the Task Group on Assessment and Testing (TGAT) is to explore the close interrelation between TGAT's terms of reference and the pilot development work on records of achievement (ROA) which we, the Records of Achievement National Steering Committee (RANSC), are steering with the objective of advising Government on the establishment by the end of the decade of national arrangements for ROA for all young people in secondary schools.
2. The background to our establishment, and the policy framework in which we are operating are set out in "Records of Achievement : A statement of policy" published by the DES and Welsh Office in 1984 and hereafter referred to as PS. We are overseeing ESG-funded pilot work in 22 LEAs in England and Wales and have established links with the other major development initiatives covering most of the remainder. We are due to report in autumn 1988 and will be issuing an interim report later this year.
3. The Government's continued commitment to this important policy initiative was re-stated in paragraph 32 of the National Curriculum consultation document (CD). In that paragraph, the proposed arrangements were described as "recording performance..... and profiling a pupil's achievement

across and beyond the national curriculum." The phrase "across and beyond" gives a clear indication that the concerns of ROA partly overlap with those of the National Curriculum (NC) and are partly separate from it. This paper concentrates primarily but not exclusively on the former.

Common aims and objectives

4. Both ROA and NC aim to raise standards. ROA seek to achieve this by giving credit for pupils' achievements, by increasing pupils' motivation, by identifying their individual curricular needs and by clarifying learning objectives (PS paragraph 11). This is fully compatible with the NC aims and purposes.
5. Though not explicitly mentioned in the PS, several of the pilot schemes have seen the need to specify clear short term objectives in pupils' learning, both to improve motivation and to make what is recorded more meaningful and precise. This practice is in line with the setting out of programmes of study and attainment targets proposed in the CD (paragraph 8(ii)).
6. ROA share with the NC the assumption that all pupils should have access to a broadly based and largely common curriculum linked to pupils' own experiences. (CD 8 (iii)). ROA are especially concerned to record those experiences which occur during or in connection with compulsory secondary schooling. These experiences were originally envisaged as important to record for their own sake; now, in the light of the NC proposals, they assume even

greater importance as a means of providing the context within which assessment against national attainment targets can be set.

7. The ROA initiative is also at one with the NC proposals in stressing the importance of continuously checking on pupils' progress throughout their schooling. ROA see the diagnostic process of stretching pupils further when they are doing well and giving them more help when they are not - CD 8(iv) - as the desirable and likely outcome of the systematic discussion and recording of pupils' achievements during the formative stages of recording (PS 35). These checks on progress - or formative assessments - are of course aimed at supporting the curriculum. We know that TGAT will be seeking to recommend assessment arrangements that similarly underpin, rather than distort, curricular objectives : the evidence from our pilot schemes as to how this can be achieved is of value here, and we will therefore be making available to TGAT our own interim report, together with the interim report from the Joint Open University/Bristol University evaluation of these projects.

8. Though the work of the pilot schemes is largely confined to the secondary phase, there are implications for the primary phase. The PS suggests that ROA should begin with a summary of pupils' experience at the end of primary education. The NC assessments at 11 would provide a valuable base for this summary, but evidence from the secondary phase suggests that these assessments would benefit from being set in the context of a broader statement of achievement. In this context, TGAT may wish to note that many teachers in our schemes would argue that the ROA formative process is one that should be begun before pupils reach secondary schools: even the age of 11 is

comparatively late to introduce such a major change in approach to learning. The NC proposals, recommending continuity from 5-16 and accompanying assessment in both primary and secondary phases, accord well with these views. We would encourage TGAT to consider the relevance of the ROA process to the primary phase and perhaps to recommend that some appropriate pilot work be undertaken.

POSSIBLE TENSIONS BETWEEN ROA AND THE NC

9. In several important respects the emphasis of ROA differs from that of the NC. In the PS and confirmed by experience in schools "it is essential that pupils be closely involved in the recording process" (PS 36). This involvement of pupils in their own assessment by means of dialogue between pupils and teachers is a key feature of ROA but a concept that is not explicitly mentioned in the CD. We commend to TGAT this key principle of pupil involvement which can both inform teachers' assessments under the national curriculum and also play a vital role in giving due emphasis to the diagnostic purposes of the assessments envisaged by the NC (CD 28).

10. Secondly, the underlying principle of ROAs is that, though both the strengths and weaknesses of pupils will be discussed during the formative process, the content of the summary document will be expressed in terms of positive achievement. If ROAs are to be the vehicle for recording and reporting NC assessments, it would be important for the results of these assessments to be in terms of varying degrees of positive achievement rather than to be of a pass/fail nature.

11. Thirdly, the NC proposals, whilst aiming to raise the standards of individual pupils, are also concerned with the accountability of schools and local education authorities. The ROA proposals are essentially for the benefit of individual pupils. This raises two possible tensions. First, certain pilot developments are strongly committed to the view that pupils are the "owners" of their ROA, having discretion about what to include and to whom to show them. The use of the summary document at age 16 - and perhaps an interim summary document at age 14 (see paragraph 17 below) to convey the results of assessments to parents would conflict with this philosophy. However most, including RANSC, would incline to the view that this principle need only be inviolate in respect to the content of the pupil's personal statement - one element of ROA. Secondly the recording and reporting of NC assessments in ways that fulfil the proposed evaluative and comparative purposes may sit uneasily within a pupil's individual record of his or her unique positive achievement, without reference to other pupils' achievements.

12. Finally, there is the question of what validation of ROA is required and/or what accreditation of schools. RANSC has been considering this difficult issue, in particular the extent to which assessment of pupils' educational attainment other than by public exam, for instance through the use of unit accreditation or staged assessment, will require external validation. In the light of the NC proposals for the moderation of assessments against attainment targets, a simpler form of accreditation/validation for ROA can now be envisaged, concerned primarily with the overall structure and processes of ROA and the personal and social elements. It will be important that the eventual arrangements for moderation

of NC assessments and validation/accreditation of ROA are complementary rather than duplicating.

Issues of Assessment

13. RANSC, like TGAT, is not primarily concerned with the content, articulation or delivery of the curriculum. Both are however concerned with its assessment and have a common interest in the resolution of a number of issues.

14. A particular concern of some pilot ROA schemes has been the assessment of cross-curricular skills and achievement. Relatively modest progress has been made to date in this field, but the issue has been given a high priority because these skills are seen by teachers as a key aspect of student success, both across the subjects of the curriculum and subsequently in the world of work. Problems of definition, and of organisation in schools which tend to be divided into single subject departments, have hampered development of what is for most anyway a challenging and unfamiliar task, and many schools have yet to develop organisational structures which can facilitate the consideration of cross-curricular issues. But ROA have the exciting potential - not least in this matter of cross-curricular skills - to give due emphasis to pupils' total achievements across the curriculum as a counterweight to any bias towards over-emphasis on performance in single subjects. Furthermore, the consultations among the staff concerned with these general learning skills - a pre-requisite, of course, for their assessment - are a most valuable step towards securing coherence between subject departments as the school's curriculum is developed.

15. Projects have been tackling in various ways the assessment of a set of objectives within subjects at various points during pupil's learning. Some projects use a unit credit approach with regular points of assessment; others are involved with staged assessment where, at least in principle, pupils are assessed when ready. The experience of schools involved in these initiatives will be of particular value and interest in relation to the NC concept of assessment and testing at predetermined ages, and we are glad that TGAT has already held discussions with some of those concerned. Equally the NC attainment targets and their associated assessments should give schools clear guidance on approaches to assessment and obviate the need for every school to devise its assessment policy from first principles. Nevertheless it will be incumbent on schools to devise a coherent policy which embraces ROA, NC and GCSE assessment in a consistent approach. To assist them, more work still needs to be done to bring together the basis for awarding GCSE grades with criteria used for internal assessments and the NC attainment targets.

Issues of recording and reporting

16. PS 38 draws attention to the need for schools to bring together with ROA as far as possible the various kinds of reports sent to parents, employers and career officers. Schools have certainly found this to be necessary in practice if the time needed for recording and reporting is to be kept to manageable proportions. We believe that this requirement should extend to the recording and reporting of NC assessments, not only for the sake of the workload in schools but also because the proliferation of different kinds of reports for parents and others could hardly be helpful.

17. The growing realisation by schools involved in ROA that they need a unified and coherent policy of assessment and recording has led to the emergence of the concept of the "interim summary" sent to parents in place of the traditional school report. The incorporation of assessment at age 14 into the appropriate interim summary should pose no difficulties other than that of avoiding negative response to which reference has already been made.

The wider purposes of ROA

18. As stated at the start, this paper has concentrated primarily on areas of mutual concern to RANSC and TGAT. It is important however to stress that in many ways the most innovative aspect of ROA is the inclusion of personal and social achievement and experience within the framework of reporting. In this very important way the aims of ROA are different from but complementary to the aims of the NC. The kinds of assessment being developed for personal achievements are very different from the nationally prescribed tests or moderated assessments envisaged as necessary for most of the national curriculum.

19. ROA, by giving weight to those aspects of the formal, informal, or extra curriculum which lie beyond the foundation subjects, can present pupils' NC achievements within the context of the totality of their achievements and experiences in and outside the classroom. They can therefore redress any tendency on the part of those who make use of the NC assessments to ascribe undue importance to achievement against predetermined subject attainment targets.

20. ROA can provide attestation of those pupils' achievements that are unforeseen but worthwhile, unplanned but fruitful. There is growing evidence that employers appreciate the importance of more general, well founded statements about young people than can be deduced from aggregation of performance in a number of school subjects.

CONCLUSION

21. The issues as complex as those relating to ROA and the NC cannot be comprehensively explored in a short paper. We would therefore recommend continuing liaison between TGAT and RANSC to further the debate and clarify the issues. We believe that this would be of benefit to both bodies.

22. At this stage, we will conclude by re-stating the most important point that RANSC wishes to register. We see it as vital that ROA and the NC develop in complementary and mutually supporting ways. The key contributions of ROA to the NC are:

- a. the formative process, with pupil teacher dialogue essential to secure the Government's objective of diagnostic NC assessments; and
- b. the setting of those NC assessments into the context of the totality of a pupil's achievements and experiences.

October 1987

TECHNICAL NOTE ON BACKGROUND FACTORS AFFECTING EDUCATIONAL PERFORMANCE

1. A strong association between social background and educational performance of almost all types is one of the longest established and best supported findings in social and educational research. Why this should be so and how pupils' social background may be measured and its effects compared remain much in dispute. Nevertheless it is now widely accepted that social background factors must be taken into account in any attempts to evaluate educational performance and effectiveness whether of individual institutions, or of LEAs. Without such analysis the effects of variations in social background are liable to conceal the effects of school, teacher or other educational variables.
2. Recent examples in this field include work by the Statistics Branch at the DES, which in a series of publications (Statistical Bulletins 16/83 and 13/84), has estimated the contribution made by various factors to variations in secondary examination performance by LEAs.
3. The study reported in SB 16/83 concluded that the two socio-economic group (SEG) variables (representing the social composition of the resident populations of whole LEA areas) accounted for the largest amount of variation between LEAs' examination results, ranging for different measures from $3/4$ to $2/3$ of the whole variation. In this study these two factors outweighed the effects of other social factors as poor housing, one parent families, unemployment, large families and free school meals. Relative levels of expenditure were even less influential in "explaining" (in the statistical rather than the educational sense) variation.
4. The most recent study confirmed that social background factors - in this case measured by the proportions of the population in high or low socio-economic groups from the 1981 Census - were easily the most important, "explaining" as much as 75% of the variation in LEAs' examination results. There was a strong association between the proportion of the population in high SEGs and the proportion in higher levels of performance (A level and 5 or more higher grades in O level/CSE), and low SEGs and lower levels, particularly the proportion of leavers with no graded result. In comparison, and in line with similar studies elsewhere, school or educational variables such as expenditure or pupil-teacher ratios "explained" very little of the variation in the LEA results.

5. The studies quoted above led to the conclusion that "any comparison between the average attainment of pupils attending secondary schools in different LEAs would be seriously misleading if it failed to take into account the social background of the pupils". Each concludes with a table comparing the actual and "fitted" or expected values, the latter representing an adjustment on the basis of the relationships between examination attainment and socio-economic factors.

6. It is possible to make these adjustments for all LEAs, because of the existence of comparable data for all areas from sources such as the national census. Similar data does not yet exist at the school level across the country. However, the DES has conducted a pilot study of the effects of social background at the level of the individual school (DES: Pilot study of school examination performance and associated factors, 1984), and some LEAs have been able to use school level data to take account of the effects of variations in intake on examination performance at secondary level. The ILEA, for example, using a similar procedure to that employed by the DES at LEA level, has presented results for individual secondary schools which take some account of variations in intake (ILEA: Looking at School Performance: Actual and Predicted Examination Scores, 1986). This analysis included data on pupil ability at intake to the secondary school, which is an important component in assessing school performance.

7. These studies at the individual school level demonstrate how data at the LEA level may tend to reduce variations by aggregating results from successful and unsuccessful schools alike. Finer grain estimates of school performance, principally based on examination results - which are not the only criteria valued by parents, employers and teachers - show considerable variations between schools, when social background and other intake differences such as ability and attainment have been taken into account. A series of research studies since the pioneering work of Rutter (Fifteen Thousand Hours, 1979) has confirmed that significant differences in the effectiveness of schools emerge once the data is examined at this detailed level (or "disaggregated" in research terms). Recent studies, for example the ILEA Junior School Study (ILEA, 1986), have suggested that individual pupil "progress" - that is change in attainment over time - may be more closely correlated with educational factors than with social background. However, social background remains the most important factors in explaining variations in performance at the group (for example school or LEA) level at any one point in time.

NUMBERS IN PRIMARY SCHOOLS

1. Statistics for England and Wales on Four Year Olds in Schools

The latest figures available for England (DES Statistical Bulletin 9/87) show that in 1986 there was a population of 1,175,000 children aged three and four. Of these 235,867 (20 per cent) were attending infant classes and 246,893 (21 per cent) were attending nursery schools or classes. (There were also 25,587 two-year-olds in nursery provision.)

These figures represent an under-estimate of the number of four-year-olds attending school, as they exclude 'rising fives' (ie those aged from four years nine months to four years eleven months). There was a slight decrease in the number of four year olds in schools in 1986 when compared with data for the previous year, but this is accounted for by a drop in the four-year-old population in 1986.

Figures supplied by the Welsh Office show that, in January 1987, there was a population of 70,600 children aged three and four. Of these 15,886 (22 per cent) were in nursery classes compared with 33,234 (47 per cent) in infant classes. (There were also 2,497 two-year-olds receiving nursery education.) There has been a tradition of admission to school before statutory age in many Welsh authorities which accounts for the high proportion of the three and four-year-old population in infant classes.

Other useful sources on this are:

- a. Results of an NFER research study Starting School at Four conducted in 1986.
- b. NFER have just published the report of a joint NFER/SCDC seminar held earlier this year. NFER (1987). Four Year Olds in School: Policy and Practice. NFER: Slough.

c. A new NFER project, aiming to research the curriculum needs of four-year-olds in school, will start in January 1988.

d. An NFER conference will be held on February 18th 1988 entitled Starting School at Four: Planning for the Future.

Reference

Department of Education and Science. (1987). Pupils Under Five Years in Each Local Education Authority in England - January 1986 - Statistical Bulletin No. 9/87.

HOW THE SYSTEM MIGHT WORK IN A PRIMARY SCHOOL

1. This appendix elaborates all the features summarised at the end of section XIII on the primary stage. It incorporates these into a more detailed account in order to convey the possible implications and consequences of our recommendations for the work of primary schools.

2. At ages 7 and 11 the basis of the reported assessments would be the combination of informal, routine assessment carried on in the classroom (enhanced by clear attainment targets and programmes of study), and of the formal standardised tasks (tests) to be set towards the end of the year. Thus the standardised tasks would constitute one part of the assessment instruments used in the primary phase. It would be possible to support the informal assessment with suitable tests set at the school's discretion; these might include diagnostic tests where the head teacher and class teachers deemed it helpful to confirm, amplify or amend judgements previously arrived at either in the informal assessment or at the formal stages of reporting. A fuller picture of a child's progress would derive from work done and assessed between ages and reported at more frequent intervals to parents.

3. We do not expect, however, that this assessment system will be additional to the many others that now exist, often side by side in the same school. We hope that the need for these - and especially for norm-referenced tests which bear little relationship to the curriculum - will disappear with the introduction of a flexible assessment system related to curriculum and progress.

Age 7 tasks

4. Towards the end of the year in which pupils become 7 the teacher would be asked to select three standardised tasks using a list of titles provided to describe topics available in a nationally prepared set of such tests. The range of tasks in the bank would enable a selection of topics to be made which would be within the children's experience, background and culture and which could be fitted unobtrusively into the run of the work being done at the time.

5. Each of these tasks would assess profile components covering knowledge, skills and understanding in numeracy, literacy, science, physical skills as well as other general skills such as observation or art. Language and mathematics, for example, might each have several profile components appropriate to age 7.

Some profile components would be identified with subjects; others would represent more general competences. Assessment for these would allow for a variety of forms of presentation, and pupils might produce responses for assessment by written work, or by oral work, or by actions which the teacher observed in a systematic way. Thus the tasks would be in keeping with primary practice, taking account of the nature and stage of young children's learning and their immediate experience while offering for the teacher clearly delineated goals and standard methods of assessing the performance produced. Pupils would be able to tackle the tasks at a variety of levels so that differentiation would be by outcome. Some tasks, or parts of tasks, might be directed at individual pupils, working on their own, others for pairs or groups. Examples of two such tasks are given in Appendix D.

6. At least one of the three tasks should be a choice agreed by a group of schools so that moderation can be undertaken on common ground. Later, when teachers have benefited from in-service training and are experienced in moderation, all three tasks might be at their choice so that they might select tasks which best suit their pupils' circumstances.

Grading

7. We expect that the work of the great majority of pupils at 7 would fall within three levels and performance would be recorded at levels 1, 2 or 3, though it would be possible for exceptional performance to rank higher than this. We envisage that most children would achieve level 2 at the age of 7. Children attaining level 1 would be those identified as needing extra help because they have made little progress. Those achieving level 3 would be children in need of work of greater pace and depth in order to use their abilities to the full. In both these cases the school would be able to make use of appropriate diagnostic tests where it considered them suitable.

Moderation

8. A teacher who has collected results for all pupils on the three tasks would be able to assign back to level 1, 2, or 3 on the basis of the results for each profile component; from these the percentage of the total who are at each of levels 1, 2, 3 could be derived. Such results could be compared with the corresponding results for the teacher's own classroom ratings, accumulated over the year. The teacher would then go to a moderation meeting where a group would exchange and compare such results and compare samples of work, in order to decide whether and how to scale the results. At the end of this, there would be an agreed overall distribution of pupils across the three levels. Subsequently the teacher could reconsider the level assignments of each of her or his own individual pupils and decide thereby on the report to the parents of each pupil.

Age 11

9. At age 11 a similar strategy would apply. Performance would be assessed using a combination of informal assessment carried out over the four years but for reporting purposes concentrating on the last year, and the formal standardised tasks set on a national scale.

10. We envisage that there would be three or four standardised tasks, again selected by the teacher from an item bank to suit the run of the work. The main difference at this stage is that the tasks would be more complex and would need to take account of the larger number of profile components which would become increasingly subject related. Although in the main differentiation would again be by outcome, it would be possible to draw on some standardised tasks offering different demands within a given profile component. There might also be more narrowly focused tests to yield information on particular profile components and so supplement the data from the task tests.

11. At 11 most pupils' performance across the profile components would fall within the levels 3, 4 and 5. All these represent progress since age 7 and we expect that most children would achieve level 5. Those obtaining level 5 (or higher) at this age would be deemed to need extra support if they are to

continue to maintain their progress. Those obtaining level 4 and below would be identified as in need of help. The use of the profile components would enable teachers to identify particular strengths and weaknesses. Any pupils who have not advanced by at least one level since age 7 would clearly need special attention.

12. As for age 7, results for each profile component would be assembled both from the national tests or tasks and from teachers' own ratings. The overall positions of pupils at each of the levels would be calculated. Teachers would then go to a group moderation meeting to discuss their results and consider action on discrepancies. The outcome would be agreed overall distributions, and an opportunity for each teacher to reconsider decisions about individual pupils before reporting to them and their parents.

Special needs

13. At both ages any child who was clearly unable to manage a standardised task (test) unaided could be helped; this help would be recorded and might signify that he or she had not attained the competences required in that profile component. If a "non-statemented" pupil was judged by the head teacher and the class teacher to be quite unable to undertake a given task for medical or emotional reasons, the child could be exempted from it and this exemption would be recorded. At both 7 and 11 children who have been "statemented" would undertake the standardised tasks to the extent specified in the statement.

Reporting arrangements

14. Our proposals for the reporting and publication of aggregated results for classes and schools are described in detail in sections XII and XIII.

HOW THE SYSTEM MIGHT WORK IN A SECONDARY SCHOOL

1. This appendix elaborates of the features summarised at the end of Section XIV on the secondary stage. It incorporates these into a more detailed account in order to convey the possible implications and consequences of our recommendations for the work of secondary schools.

Entry to secondary school

2. Under the national assessment system, all secondary schools would be receiving criterion-referenced and calibrated results based on the collection of data for every pupil on transfer from their feeder schools. It should be clear, from this transfer assessment information, what pupils know, understand and can do after the primary phase of education; and this will be directly related to the attainment targets set for core and other foundation subjects. This information should form a sound basis for whole school and departmental decisions about the learning needs of the new intake of pupils, so that continuity and progression are more likely to be achieved. At least some of the transfer information would be specific to subjects and profile components, so that relevant departments would be able to allocate tasks at the appropriate level of challenge for pupils from the outset of the secondary phase. Teachers would also have indications of the likely range of pupils' future performance to build on at the transfer stage, based on the primary school's assessment of the general knowledge, understanding and skills which have applicability across the whole curriculum at secondary schools.

Course planning

3. Overall course planning and directly-related pupil activities for class and homework, would need to ensure that all the specified subject profile components were covered in a balanced and relevant way throughout the programme of study. Although the system allows for assessment of single profile components to illustrate varying pupil strengths and weaknesses, the overall characteristics

and cross-curricular relevance of subjects would need due attention in teaching and assessment practice too. Good practice should also ensure that all teaching groups were properly informed about the relevance and meaning of subject profile components and their assessment, so that pupils were encouraged to be in a positive partnership at classroom level aimed at clearly understood and shared learning and assessment objectives.

Ages 11 to 14

4. By the age of 14, as earlier, the assessment and its recording would derive from a combination of routine teacher assessment of normal class work and homework against clear objectives, with inter-class and inter-school group moderation of selected work samples, and the use of external tests. All this information would determine the levels of achievement reached by each pupil in the profile components of the foundation subjects. The teacher assessment evidence will be built up and refined over the three years from 11 to 14, though concentrated for reporting purposes in the last of these years. The national tests would be used near the end of the age 14 year. The existence of subject departments would allow for sharing of both the formal and informal work of assessing and recording. Through departmental meetings the assessments, with samples of pupils' work, could be moderated across classes and an overview of the levels reached by each pupil could be built, in terms of the level derived from the teacher's own assessment and the corresponding level results from the national tests. As for age 11, these results would then be prepared and taken to an inter-school moderation meeting. Each such meeting would now be concerned with one of the foundation subjects. The aims and outcomes would otherwise be similar to those for age 11.

5. We expect that at age 14 most pupils who have made progress since 11 would achieve levels in the range 5 to 7, though the span of performance at this stage could be reflected in the award of levels outside that range. The levels for national assessment would be related to GCSE in the sense that achievement within their normal range at age 14 should indicate that, given continued progress, the pupil was on course for grades at GCSE. The review of progress might be aided by reference to each pupil's results at age 11.

6. In the intervening years between 11 and 14, schools might wish to use not only their own curriculum-related tests (whether of their own devising or those associated with the particular courses), but also diagnostic tests - whether nationally provided or commercially available - for cases where specific needs have to be identified. This process should be assisted by the possibility of comparing with results, at age 11, both for individual pupils and for the school as a whole.

7. Levels of progression defined within each profile component should enable teaching practice to aim for appropriate differentiation of tasks and outcome expectations directly related to the pupil stage of achievements, so that sufficiently challenging tasks are set for progression. The shorter-term goals in stages of achievement should also help to promote greater self-awareness of progress and encourage pupil motivation.

8. The use of records of achievement as a vehicle for recording results would mean that the information was set out in the context of general development, giving credit beyond the individual profile components.

Age 16

9. At 16 the great majority of pupils will be entered for GCSE and other nationally approved examinations which are available for certification and qualification purposes. These examinations would however need to be supplemented by national assessment for those pupils not entered in particular subjects, and especially for those who leave school at Easter. As at age 14 the basis for the assessment would be the routine assessment of pupils' work, building on that at 14 and supported by internal and external moderation; standardised tasks (tests) would also form part of the assessment at this stage. The levels would, as at earlier ages, relate to profile components; but for the reasons explained in sections XI and XIV of the report we do not envisage at this stage a precise link with to GCSE grades. The results would be incorporated in each pupils Record of Achievement and the national assessment procedures should thus help to maintain it even where a pupil is not entered for GCSE.

Reporting

10. At ages 14 and 16 the results of the assessments would be made available in detail to pupils and their parents, and performance would be set against aggregated national results. At 16 the reporting would be largely in terms of GCSE and other examinations. Publication of aggregated results is dealt with in detail in section XII.

Effects on teaching and learning

11. We expect that methods of national assessment will, with the national curriculum changes, have beneficial effects on styles of teaching and learning. The requirement that individual pupil progress be recognised and encouraged will have teaching implications across departments, calling for a more flexible approach to the processes of learning. There is likely to be an increasing need to develop active learning utilising a range of methods beyond the didactic and into discovery, experiential and problem solving approaches which allow for and encourage individual pupils' progress and state of readiness. In particular, teaching practice will need to give due weight to the levels of achievement which pupils have reached, through the selection of tasks set by the teacher, for example by use of open ended investigation to complement the more structured task with teacher guidance. In practical situations, including but not confined to science, adequate non-teaching support and proper use of technician staff in preparatory work and lesson organisation will be increasingly important so that all levels of ability and progress are properly catered for in the relevant subjects.

RESOURCE NEEDS - SOME DATA ON EXISTING AND PLANNED PROVISION

These data are presented here as a general background. No attempt is being made for this report to draw conclusions directly bearing on the costs of a new national assessment scheme.

1. LEA Training Grants Scheme

In the 1988/89 arrangements for LEA Training Grants Scheme, provision is made for grant-related expenditure of £15m on training in the assessment of achievement. Training which is eligible for grant aid under this heading comprises any training for primary, secondary and FE teachers which is concerned mainly with the development of skills required to "assess, record and report achievement". It is clear from the commentary in Circular 9/87 that most of this expenditure is aimed at secondary work and in particular is geared to development of these skills in the GCSE context. It is worth noting that this £15m is part of a total of £77m identified under various national priority headings (on which grant is payable at a rate of 70%). In addition the 1988/89 programme provides for grant-related expenditure of up to £130m on locally assessed needs (on which grant is payable at a rate of 50%).

2. Education Support Grants: Records of Achievement

Some developmental work is being undertaken with ESG funding through the Records of Achievement project. In England eight development schemes involving 14 LEAs are currently being supported at a cost of £2.25 million per annum. The grants were introduced in 1985/86 and will cease at the end of the 1987/88 financial year. An extension of grant for a further two years will be offered for 5-8 pilot schemes, involving expenditure of not more than £1.2 million in 1988/89 and a similar amount in 1989/90. In addition a further pilot scheme is running in Wales and covers all eight Welsh LEAs.

3. Teachers Conditions of Employment

The new Conditions of Employment for Primary and Secondary Teachers include as part of the professional duties which a teacher may be required to perform, the assessing, recording and reporting on the development, progress and attainment of pupils. They also include "participating in arrangements for preparing pupils for public examination and in assessing pupils for the purposes of such examinations; recording and reporting such assessments; and participating in arrangements for pupils presentation for and supervision during such examinations". The conditions cover the number of days (195) and the number of hours (1265) in a year for which a teacher may be required to work at specific times in a prescribed place.

As far as the 195 days are concerned, five of these are days on which pupils are not present and which may be used for, among other things, INSET activities. Some schools/authorities may be using most of these five days for INSET: the general experience seems to be that two or three of them are available for this purpose. In many cases these additional days would only become available at the beginning or end of term and there could well be logistical problems in using these days for a major INSET exercise at LEA level, given the large numbers of teachers involved.

4. The cost of a day's INSET

The cost of providing a day's INSET for a teacher will vary from authority to authority depending upon local policies relating to the reimbursement of travel and subsistence expenditure incurred by teachers in attending courses, and upon the arrangements whereby authorities deliver their INSET programmes. The full cost could be as high as £75 per day, on average might be nearer a figure of £50 and in some cases could obviously be below this figure. The size of the teacher force nationally is of the order of 170,000 in primary schools and 235,000 in secondary schools (DES 1985 statistics). On an average basis it would require something over £20m to provide one day's INSET for every teacher where this involved absence from teaching and the provision of supply cover.

5. External Bodies

At present, substantial expenditure is incurred on 16+ examinations through GCSE Boards and certain other examining bodies. Most of this expenditure is derived from candidates fees for which LEAs are mainly responsible.

There are also two other bodies (the APU and NFER) which are involved in assessment/testing activity. The APU currently operates on a budget of about flm annually and the NFER on one of about f2.5m.

6. Education Reform Bill - Financial Memorandum

The Explanatory and Financial Memorandum to the Education Reform Bill contains the following paragraph:-

"The implementation of the National Curriculum will lead to estimated additional public expenditure rising to f33 million in 1990-91. This covers the establishment and activities of the National Curriculum Council and the School Examinations and Assessment Council (clause 7), an element for research and development particularly in the areas of assessment and testing, and the administration of assessment arrangements, for which local authorities will not be charged fees. As the National Curriculum is progressively introduced, some redirection of recurrent and capital resources will be required by local authorities, school governing bodies and headteachers as well as by central government in its planned development of education support and training grants."

7. Other Data

Relevant data which might be considered with the above are the costs of development work of the examining groups for GCSE, and the resource used by local authorities, aided in many cases by grants from central government, to establish the graded assessment and records of achievement systems. It may be helpful to consider these, with the above, when consideration of the system needed to support national assessment, is further advanced.